Sequencher®
User Manual

The complete software solution for sequencing DNA

Windows

Gene Codes Corporation

THIS IS AN AGREEMENT BETWEEN YOU, THE END USER, AND US, GENE CODES. BY OPENING THE SEALED DISK PACKAGE, YOU ARE AGREEING TO BE BOUND BY THE TERMS OF THIS AGREEMENT. IF YOU DO NOT AGREE TO THESE TERMS, PROMPTLY RETURN THE UNOPENED DISK PACKAGE AND ALL ACCOMPANYING MATERIALS TO US OR TO YOUR DEALER FOR A REFUND.

License. We grant you the right to use the enclosed software program (the Software) on a single terminal connected to a single computer. The Software is accompanied by a hardware-locking device that prevents its full use on any computer without the device; if the Software is used without the locking device, it operates in demonstration mode only. You may not use the Software on a network without obtaining a license for each terminal on which the Software will be used. You may not use the Software for any purpose related to the use, development, analysis, testing, or research on biological weapons.
Copyright. The Software is protected by United States copyright laws and international treaty provisions. You must treat the Software like any other copyrighted material, such as a book or musical recording. You may, however, either
(a) make copies of the Software for backup or archival purposes; (b) transfer the Software to a single hard disk, provided you keep the original; or (c) make copies of the Software to give to others for purposes of demonstrating the Software. You may not copy any written materials accompanying the Software.
Other restrictions. You may not rent or lease the Software, but you may transfer it and the accompanying written materials on a permanent basis provided you retain no copies and the recipient agrees to the terms of this agreement. You may not reverse engineer, decompile, or disassemble the Software.
Limited Warranty. We warrant that the Software will perform substantially in accordance with the accompanying written materials for a period of 90 days from the date of receipt. Any implied warranties on the Software are limited to 90 days. Some states do not allow limitations on duration of an implied warranty, so this limitation may not apply to you.
Remedies. Our entire liability and your exclusive remedy shall be, at our option, either (a) return of the price paid or (b) repair or replacement of the Software that does not meet our limited warranty and that is returned to us. Our limited warranty is void if failure of the Software has resulted from abuse or misapplication. We warrant the replacement Software under this limited warranty for the remainder of the original warranty period or 30 days, whichever is longer.

NO OTHER WARRANTIES. WE DISCLAIM ALL OTHER EXPRESS OR IMPLIED WARRANTIES, INCLUDING THE IMPLIED WARRANTIES OF THE MERCHANTIBILITY AND FITNESS FOR A PARTICULAR PURPOSE. OUR LIMITED WARRANTY GIVES YOU SPECIFIC LEGAL RIGHTS. YOU MAY HAVE OTHER RIGHTS, WHICH VARY FROM STATE TO STATE.
NO LIABILITY FOR CONSEQUENTIAL DAMAGES. IN NO EVENT SHALL WE BE LIABLE FOR ANY DAMAGES WHATEVER, INCLUDING LOSS OF PROFITS, ARISING OUT OF THE USE OR INABILITY TO USE THIS SOFTWARE, EVEN IF WE HAVE BEEN ADVISED IN ADVANCE OF THE POSSIBILITY OF SUCH DAMAGES. BECAUSE SOME STATES DO NOT ALLOW THE EXCLUSION OR LIMITATION OF LIABILITY FOR CONSEQUENTIAL OR INCIDENTAL DAMAGES, THIS LIMITATION MAY NOT APPLY TO YOU.

U.S. Government Rights. The Software and documentation are provided with RESTRICTED RIGHTS. Use of duplication, or disclosure by the Government is subject to restrictions set forth in subparagraph (c) (1) (ii) of the Rights in Technical Data and Computer Software clause at 52.227.7013. The contractor/ manufacturer is Gene Codes Corporation, Ann Arbor, Michigan.

© 2011 Gene Codes Corporation, Inc. All rights reserved

Sequencher® is a trademark of Gene Codes Corporation

Microsoft® is a registered trademark of Microsoft Corporation.

Macintosh® is a trademark of Apple Computer, Inc.

ABI, Applied Biosystems and 3730 are trademarks of Applied Biosystems, Inc.

ALF is a trademark of Pharmacia Biotech.

All other trademarks are the properties of their respective owners.

1. PREFACE 1

Gene Codes 1

2. USING THIS MANUAL 3

Conventions used in this manual 3

Installing Sequencher 3

Sequencher Help 3
- Navigating Sequencher Help 4
- Context Sensitive Screens 4

The button bar 5

Sequencher's menus 5
- Sequencher 5
- File menu 5
- Edit menu 5
- Select menu 5
- Contig menu 6
- Sequence menu 6
- View menu 6
- Window menu 6
- Help menu 6

Context Sensitive menus and buttons 6

Contextual menus 6

Keyboard shortcuts 7

3. THE PROJECT WINDOW 8

Sequencher concepts 8
- About the Project Window 8
- Creating a new project 8
- Opening an existing project 9
- Working with items in the Project Window 9
- Icon types 10

Project views 10
- View as Large Icons 10
- View as Small Icons 11
- Moving icons 11
- Cleaning up 11

View as a list 12
Project Window Columns 12
Viewing constituent sequences 13
Viewing the information for a sequence or contig 14
Editing the information for a sequence or contig 15
The Project Window button bar 15

Templates 15
Save Project As Template 16
New Project From Template 16
Import from Template 16

4. IMPORTING DATA 17

Quick and easy importing of data 17
Dragging and dropping files 17
Copy and paste 17

How to import using menu commands 18
Importing sequences 18
Importing GenBank features 18
Importing lists of sequences 18
Folder of sequences 19
Projects in specific formats 20
Sequencher projects 20
Files from VecBase 20
Creating a new sequence 21
Double-clicking a project icon 21
Importing Confidence Scores 22
Importing PHRED called data 22
Importing PHRAP Files 22
Removing sequences and contigs from a project 23
Closing an editor window 23
Closing a project 23

5. ORGANIZING PROJECT VIEWS 25

Working with icons and lists 25
Changing between project views 25
Collecting sequences in refrigerators 25

Selecting sequences 26
Selecting sequences and contigs 26
Selecting with the mouse 26
Selecting by typing 26
Selecting multiple items with a marquee 26

Selecting using menu commands 27
Selecting all icons 27
Selecting all items containing sequence 27
Finding assembled sequences 27

Renaming a sequence or contig 28
Renaming using the mouse 28
Renaming using the menu 28
Labeling items in the Project Window 28

Saving your work 29
Saving your Project 29
Auto-Save 29
Saving a duplicate of a current project 29
Reverting to the saved version of the project. 29
Closing the Project Window 30
Exporting data 30
Exiting the program 30

6. PREPARING YOUR DATA FOR ASSEMBLY 31

Removing undesirable data 31

Trimming poor quality data 31
Selecting and trimming fragment data 31
Performing a trim with default settings 32
Automatically selecting sequences for ends trimming 32
How to set the Ends Trimming criteria 32
Using Confidence to Trim Ends 33
Reviewing the Trim 33
Sorting items 34
Executing the trim 34

Trimming vector contamination 35
Selecting sequences for Vector Trimming 35
Specifying a vector using VecBase 35
Specifying the insertion site manually 37
Saving and loading vector sites 37
Setting Vector Trimming Criteria 38
How Sequencher screens sequences 39
Executing the trim command 39

7. THE REFERENCE SEQUENCE 40

What is a Reference Sequence? 40
Why use a Reference Sequence? 40

Reference Sequence properties 40
Basic Reference Sequence properties 41

Working with a Reference Sequence 42
How to mark a sequence as a Reference 42
Assemble to Reference 42
To Reference by Name 42
Trim to Reference 43
Contig Editing with a Reference Sequence 43
Fill Coverage Gaps with Reference 44

Reference Sequence Translation 44

8. THE SEQUENCE EDITOR 45

Opening an existing sequence for editing 45
Locked Editors 45
Basic sequence editing 46
Setting the base numbering 47
Set Circular Genome Size 47
Splitting a sequence 47
Duplicating a sequence 48
Reverse and Complement 48

About experimental data 48
Creating a baseline 48
Resetting a baseline 49
Viewing experimental data 49
Reverting to experimental data 49
Viewing experimental data for Chromatograms 49
Viewing summary information in the Sequence Editor 50

Selecting Bases 51
Finding Ambiguous Bases 51
Finding Open Reading Frames (ORF) 51

Sequence features 52
Looking at codon maps 52
Looking at Restriction Maps 52
Annotating Sequences 53

Customizing your sequence view 54
Formatting Ruler 54
Margins 54
Translation 55
Base grouping 55
Line spacing 55
Case and Font 55

Voice verification 56
Selecting a voice file 56
Identifying a speaker 56
Audible keystrokes 56
Read sequence selection 56

9. SEQUENCE ASSEMBLY 58

The Assembly Strategy 58

Setting the Assembly Conditions 59
Setting assembly parameters 59
Assembly Algorithms 59
Assembling with Dirty Data 60
Assembling with Clean Data 60

Assembling with Large Gap 61

Refining the Assembly conditions 62
Minimum Match Percentage 62
Minimum Overlap 62
Maximum Loop Out Size 62
Optimization of gap placement 62

Performing the assembly 63
Automatic assembly 63
Adding selected items to others – incremental building of contigs 63
Assemble to Reference 63
Interactive assembly 64
Performing an interactive assembly 64
Changing parameters for an interactive assembly 67
Mindlessly Join 67

10. ASSEMBLE BY NAME 68

Setting the Assemble by Name Conditions 69
Configuring Assembly Handles using a Single Delimiter 69
Configuring Assembly Handles using Advanced Expressions 70
What is a regular expression? 70
Setting your Assembly Handle Names 71
Choosing a New Assembly Handle 72
Enabling Assemble by Name 72
Setting Assembly Parameters 73

Performing an Assembly with Assemble by Name 73

11. THE CONTIG EDITOR 75

Contig editing 75

Overview of the contig 75
The contig overview window 75
Selection marquee 76
Checking coverage in the Overview 76

Overview features 77
Bases 77
Summary 77
Sort 77

Overview options 77
Scale Diagram To Window Size 78
Diagram Key 78
Labels Of Fragments 78
Names Of Fragments 78
Names Of Fragments, At Left 78
Positions Of Fragments 79
Start & Stop Codons 79

Base Numbers At Transitions 80
Base Numbers At Every x Bases 80
Condensed Base Numbers 80

Getting more information 80
Find 80
Get Info 80

The Bases View 80
Working in the Bases View 80
Viewing sequence names 82

Before you start to edit 82
Setting the base numbering 82
Consensus calculation 82
Consensus Inclusively 83
Consensus By Plurality 83
Consensus to Forensic Standards 84
Assigning a Reference Sequence 84

12. EDITING CONTIGS 85

Finding bases which need attention 85
Viewing an individual sequence from the contig 85
Finding disagreeing ambiguities 85
Finding Disagreeing Bases 86
Finding Low Confidence Bases 86
Finding Edited Bases 86

Performing edits 86
The Edit Command 86
Editing modes 87
Viewing Base Edits 87
Collect Gaps 87
Moving Bases and Gaps 88
Using ReAligner to clean up contigs 88
Moving sequences 89
Deleting bases 89
Deleting bases from the 5' or 3' ends 90
Inserting gaps or bases into a contig 90
Making edits from the consensus line 90
Split After Selection… 91
Auto Match 91
Create new sequence from a consensus 92
Removing sequences from a contig 93
Dissolving a contig 93

13. THE VARIANCE TABLE 94

The Variance Table Strategy 94
What is a Variance Table 94
The Variance Table and consensus sequences 94
The Translated Variance Table 95

The Structure of the Variance Table 95
Structure of the Translated Variance Table 97

Creating a Variance Table 98
Variance Table for sequences in the same contig 98
The consensus Variance Table 98
Translated Variance Table for sequences in the same contig 99
The Translated consensus Variance Table 100

Setting the Variance Table Conditions 101
Restricting the Comparison Range 101
Restricting the Translation Range 102
Setting the Variance Table Display Options 105
Setting the Consensus Calculation for a Variance Table 106
Variance Table Options 107

Working with the Variance Table 107
Navigating a Variance Table 107
Sorting the Variance Table 107
Removing columns from the Variance Table 108
Review Mode 108
The Review Mode and Compare Consensus to Reference 109
The Review Mode and the Translated Variance Table 110
The Review Mode and Sister Tables 111

Variance Table Reports 112
Variance Table Report 112
Individual Variance Reports 113
Variance Detail Report 114
Population Report 114

Working with Variance Table Reports 115
Printing a Variance Table Report 115
Printing selected data from a Variance Table 116
Exporting a Variance Table Report 118
Exporting selected sample sequences 118
Exporting data for selected variants 119

Translated Variance Table Report 120
Translated Variance Table Report 120

Working with a Translated Variance Table Report 120
Exporting a Translated Variance Table Report 120
Exporting selected sample sequences 121
Exporting data for selected variants 122
Removing data from a table before export 123

14. THE SUMMARY REPORT 124

The Summary Report view 124
View by Summary 124

Summary report display options 125

Compare To Consensus or Reference Sequence 125
Icons 125
Bullets, Pluses and Dashes 126
Consensus Sequence 126
Fragment Sequences 126
Matching Bases as Dashes 126

Protein translations and the Summary Report 126
Enabling the protein display 127
Protein translations for the Consensus or Reference Sequence 127
Protein translations for Fragments 127
Matching Proteins As Dashes 127
Using the formatting ruler 128
Reverse and Complement a contig 128

15. CHROMATOGRAMS 129

Working with automated sequencer data 129
Viewing chromatograms from a Sequence Editor 129
Viewing chromatograms from a contig editor 130
Understanding the Trace Display 131
Display secondary peak 132
Select Next Commands 133
Editing bases from traces 134
Reverting to experimental data 134
Reverting assembled sequences back to experimental data 135
Adjusting trace positions 136

16. RESTRICTION MAPS 137

Displaying a Restriction Map 137

Restriction Map display options 137
How to select enzymes 137
How to show cut positions or fragment sizes 138
How to set the map style 138
Setting the map width 139
Setting the map caption 139
Getting more information about an enzyme 140
Changing the selected enzyme 140
Changing the default enzymes 142
Adding and editing enzymes 142
Changing the recognition sequence 143
Salt concentration effects 143
Saving new or modified enzymes 143
Setting the sequence selection in a Sequence Editor 143
Setting the sequence selection in a contig editor 143
Copying the map 144

17. MOTIFS AND FEATURES 145

Motifs 145

Entering motifs 145
Quick motif entry 146
Displaying motifs 146

Features 147
What is a feature 147
Creating and editing features 148
Editing an existing feature 149
Quick feature creation 149
Showing and hiding features 150
Listing a sequence's features 150

18. FINDING ITEMS 151

Finding the Project Window 151
Finding open windows with menu commands 151
Finding open windows in the Project View 151
Finding the current selection 152

Searching in Open Windows 152
Finding a subsequence 152

Refining a subsequence search 153
Exact matching 153
Matching ambiguous bases 153
Any ambiguous match 153
Finding bases by number 153
Extend Selection 153
Other search options 153

Project Window selection 153
Finding an item by name 153
Selecting all items that… 154
Contain a subsequence 155
Contain Items Named 155
Contain Items With Names Containing 155
Have Chromatograms 156
Have Comments Containing… 156
Have Labels Containing… 156
Have Names Containing… 157
Were Edited On Or Since… 157

19. EXPORTING DATA 158

Exporting data from Sequencher 158
Exporting Sequences 158
Exporting a Consensus 159
Exporting Contigs 160
Exporting Selection as subproject 160
Export Selected Bases 161
Export Options 161
Export Formats 161
Exporting protein translations 162

Export Overviews 163
Exporting contig summaries 163

20. OUTPUT 164

Copy picture 164
Print setup 164
Print Trace in One Page 164
Printing in detail 164
Setting header, footer and margin options 164
Report formats 165
Page breaks 165

21. CUSTOMIZING SEQUENCHER AND USER PREFERENCES 167

View Preferences 167
Ambiguous bases 167
Edited Bases 167
Display Color Bases 167
Colors as Backgrounds 168
Display Base Confidences 168
Display Features 168
Display Motifs 168

Setting up custom codes 168
Standard coding system 168
Configuring Ambiguity Codes 168
Replacing an existing character keystroke code 169
Saving and loading custom codes 169
Show and hide ambiguity helper 169
Choosing a genetic code 170
Editing a genetic code 170
Abbreviations 170
Remember Window Layout 171

User Preferences 172

General 172
Settings 172
Auto-Save 172
Confidence 173
Label & Name 174
Menu 174
Sound 175
New Project 175

Display 176
Chromatogram 176
Contig Chromatogram 177
Contig 178
Feature, Motif 179
Format Ruler 180
Start-Stop 181

Variance Table 182

Input/Output 182
File Import 182
Report 183

22. FORENSIC FEATURES 184

Working with mtDNA profiles 184
Validating mitochondrial DNA profiles 184
Create Reports 185
Export CMF 185

Further Commands of Interest 186
Set Circular Genome Size 186
How to mark a sequence as a Reference Sequence 187
New Project From Template 187
Display secondary peak 187

23. APPENDIX - KEYBOARD SHORTCUTS 189

24. APPENDIX – ADVANCED EXPRSSIONS 191

25. APPENDIX - IUPAC-IUB STANDARD CODE 193

26. APPENDIX - STANFORD EXTENDED CODE 194

27. APPENDIX – DEFAULT FEATURE STYLES 196

28. APPENDIX - FEATURE KEYS AND QUALIFIERS 199

GenBank Feature tables 199
Feature table qualifiers 201

29. APPENDIX - DEFAULT FEATURE QUALIFIERS 203

30. GLOSSARY 206

1. Preface

Sequencher is the premier DNA software choice for DNA sequence assembly and analysis. Its capabilities include:

- Multiple, configurable DNA assembly algorithms
- Comprehensive DNA sequence editing tools
- Full support of sequence data confidence values
- Powerful Reference Sequence and Variance Table to find SNPs quickly and easily
- Restriction Mapping
- Extensive data import & export capabilities, including customizable GenBank Feature handling
- Specialized tools for Forensic mtDNA analysis
- Comprehensive HTML Help throughout

Almost 20 years of daily use by biologists in labs around the world have refined Sequencher's tools and interface. You get the power and the speed to get the best, most accurate results from your DNA analysis, and get back into the lab more quickly.

Gene Codes

Gene Codes is a biotechnology-oriented software developer whose headquarters are in Ann Arbor, Michigan. Our goal is to write powerful software tools that are easy to use. If you have any suggestions on how to improve this product, please contact us. You can telephone us at (734) 769-7249, fax us at (734) 769-7074, or write to us at:

Gene Codes Corporation
775 Technology Drive, Suite 100A
Ann Arbor Michigan 48108
http://www.genecodes.com

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

We welcome your comments or support questions by electronic mail at:

support@genecodes.com

We welcome your general inquiries by electronic mail at:

gcinfo@genecodes.com

2. Using this manual

This chapter will introduce you to the conventions we will be using in this manual and how to get help. We will then introduce you to the button bar and Sequencher's menus, which offer you a rich range of options while maintaining a scientist-friendly interface.

Conventions used in this manual

This manual follows certain conventions for displaying text and pictures and for alerting you to special information.

Installing Sequencher

Sequencher is easy to install. Your installation package will include a CD, a Sequencher key (dongle), and an installation guide. Follow the instructions in the installation guide to install your software. You must have Administrator privileges in order to install Sequencher.

Sequencher Help

Sequencher offers help through a Help viewer. To display Help, go to the **Help** menu and choose **Sequencher Help** or press **F1.**

Sequencher Help is organized into two main areas: application screens and the Sequencher User Manual. The Sequencher application screens contain context sensitive links to relevant sections of Sequencher Help.

Note: You can leave the **Help** window open and return to Sequencher by clicking on an open window.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Navigating Sequencher Help

The Sequencher Help window consists of three main parts. There is a button bar at the top of the window. There is also a navigation pane on the left side of the window and a content pane on the right side of the window.

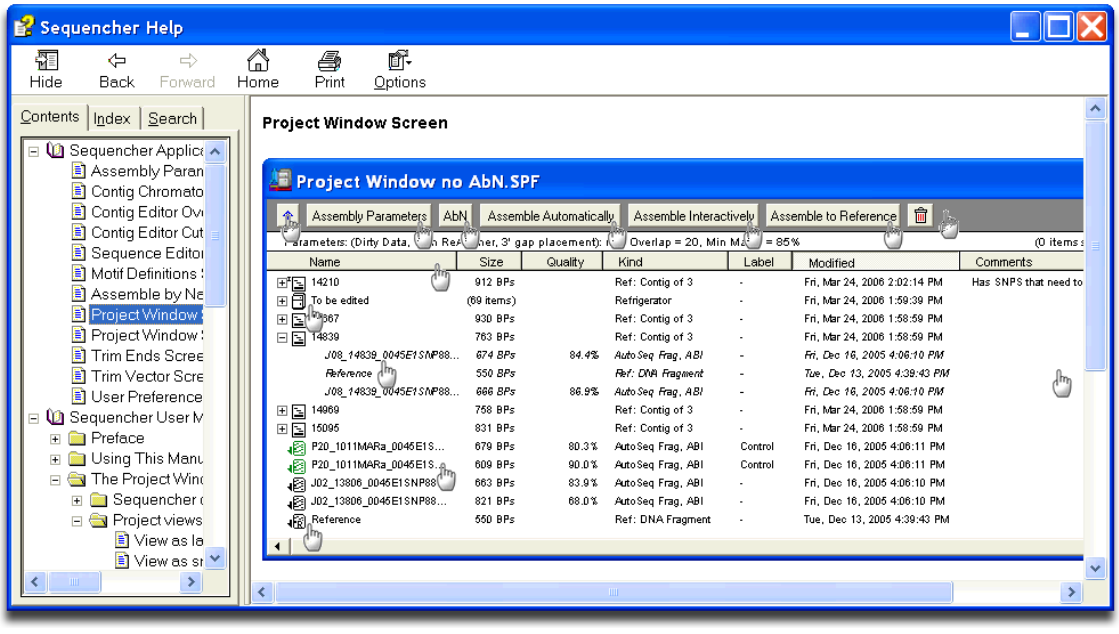
The button bar has **Back** and **Forward** icons which function like a web browser's Back and Forward buttons. If you click the **Home** button you will return to the **Welcome** screen. You can also type keywords into the **Search** entry text box.

The navigation pane displays a Table of Contents. Click on the relevant icon to display any subtopics. Click on a page link icon in the navigation pane to display the Sequencher Help article in the content pane. Some of these pages contain underlined hyperlinks to related topics in the manual and a hyperlink to the following topic in the Table of Contents.

You can also get information on specific topics such as **Assembly Parameters** or the **Sequence Editor** from context sensitive screens.

To search for a specific topic, type the topic name into the search field at the top right corner of the Help Viewer. When your search is complete, double click on an item to display it.

Figure 2-1 The Sequencher Help dialog window



Context Sensitive Screens

Context sensitive screens provide easy links to information about the region of Sequencher you are currently using. If you have the Project Window displayed when

accessing Help, then the Project Window context page will appear. This will be the case for most of Sequencher's major windows.

If you see a flagged icon, this means you can click on a link to get further information. If you place the cursor over this icon, a Tool Tip with the title of the Help article will be displayed.

The button bar

The button bar contains buttons for frequently used commands. It appears just below the top of many windows (see Figure 2-2).

Figure 2-2 The Project Window button bar

On many Sequencher windows you can use the smaller left-hand button, which has an up arrow on it, to hide the button bar. Click on the button and the button bar will "hide" behind the title bar.

When the button bar is hidden, you can still see the bottom of the left-hand button under the title bar. It will show a down-pointing arrowhead. Click on the arrowhead and the button bar will reappear.

Sequencher's menus

Sequencher

The Sequencher menu gives you information on the version of Sequencher you have, contact information for **Gene Codes Corporation** and **Sequencher** technical support, access to the **User Preferences** and the **Quit Sequencher** command.

File menu

The commands in the **File** menu let you open, close and save projects as well as import and export data. You can also open and close windows, specify page setups and print from the **File** menu.

Edit menu

The commands in the **Edit** menu let you cut, copy and paste selected data. You can also duplicate fragments, mark motifs, reposition sequences and choose your editing mode from the **Edit** menu.

Select menu

The commands in the **Select** menu cover a variety of items you might want to find and/or highlight to help you in your analysis. These commands use a wide variety of criteria to help you locate items from sequence fragments, subsequences and individual bases.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Contig menu

The commands in the **Contig** menu facilitate most of your work in contig editing. You can set assembly parameters and consensus mode, assemble, dissolve and rename your contigs, reposition sequences, remove selected sequences and **Fill Coverage Gaps with the Reference Sequence** or **Trim to Reference Sequence**.

Sequence menu

The commands in the **Sequence** menu facilitate most of your work in sequence editing. You can set up data entry preferences such as trimming poor quality data or vector contamination, edit features, create and rename sequences, set up a Reference Sequence, set up base numbering and **Revert To Experimental Data.**

View menu

The commands in the **View** menu let you control how data is displayed as you work on it. You can add extra information to the display, specify the format and marking of the data and organize project files from the **View** menu. Some of the functions in the **View** menu have options, which you can change by going to **User Preferences** (see Chapter 21 "Customizing Sequencher and User Preferences", for more information).

Window menu

The commands in the **Window** menu offer access to different kinds of user help, to reference information on various aspects of your data, to different windows you have opened and to **User Preferences**.

Help menu

The command in this menu offers access to user help.

Context Sensitive menus and buttons

Sequencher has a number of menus and buttons that may change depending on the context. Some items will become enabled. The labels of other items will change. This feature is designed so that Sequencher can offer you extensive options within the same easy-to-use interface.

Contextual menus

Most windows in Sequencher offer contextual menus. You can invoke these by clicking within the window you are reading while holding down the right hand mouse button. If you select one or more items and then click the right hand mouse button, you will see other menus.

Note: If you have a mouse with only one button, you may be able to mimic the effect of a right hand button with the use of a modifier key.

Keyboard shortcuts

Sequencher has a number of keyboard shortcuts for major menu items and other important features and allows you to define some of your own. See Appendix 23 “Keyboard Shortcuts” for a complete listing. See the Chapter 21 on “Customizing Sequencher and User Preferences” for more information about how to define your keyboard shortcuts.

3. The Project Window

In this chapter we will introduce you to Sequencher's Project Window. We will also show you how to get started with Sequencher by creating a new project or opening an existing one. This chapter also describes how you view the constituent sequences or contigs in your project and how to annotate information in your project for later editing and record keeping.

Sequencher concepts

About the Project Window

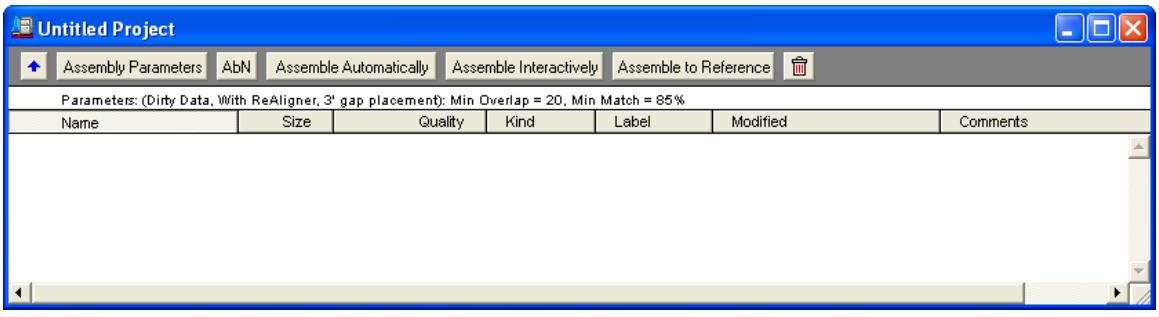
The concept of the project is central to Sequencher. Users work within the framework of a project. A Sequencher project comprises a collection of DNA sequences and contigs (contiguous alignments of overlapping sequences) that are built from those sequences. A project can be as large or as small as you want.

Sequencher stores the sequences you enter, the information on how these sequences fit together to create any contigs you have formed from them and information on user-specified parameters that control the alignment operations, all combined into a single data file. The Project Window displays all of your sequences and contigs.

Creating a new project

When you launch Sequencher by double-clicking the program icon, a new, empty project is created (see Figure 3-1).

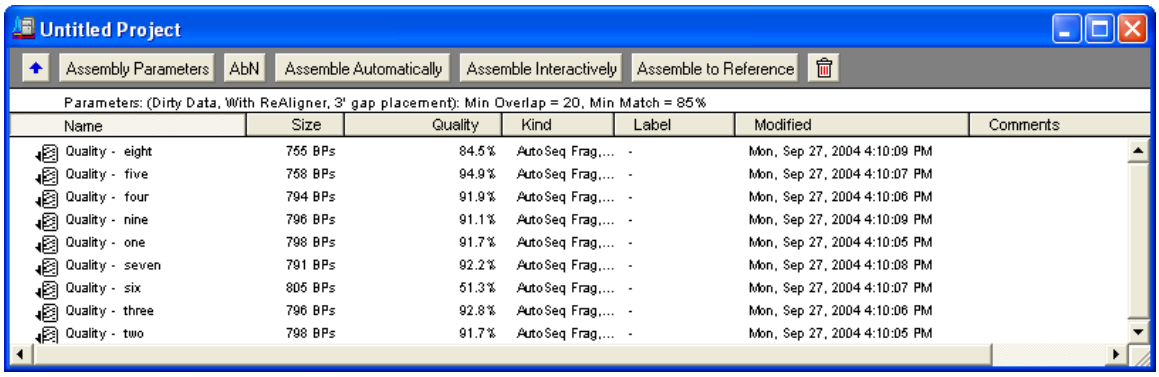
Figure 3-1 An empty Sequencher Project Window



SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Until you create or import sequence fragments, the new project will remain empty. As you add sequences to your project and build contigs from them, the Project Window begins to look more like Figure 3-2.

Figure 3-2 A new project


Untitled Project
Assembly Parameters
AbN
Assemble Automatically
Assemble Interactively
Assemble to Reference
Parameters: (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20, Min Match = 85%
Name
Size
Quality
Kind
Label
Modified
Comments
Quality - eight 755 BPs 84.5% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:09 PM
Quality - five 758 BPs 94.9% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:07 PM
Quality - four 794 BPs 91.9% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:06 PM
Quality - nine 796 BPs 91.1% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:09 PM
Quality - one 798 BPs 91.7% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:05 PM
Quality - seven 791 BPs 92.2% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:08 PM
Quality - six 805 BPs 51.3% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:07 PM
Quality - three 796 BPs 92.8% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:06 PM
Quality - two 798 BPs 91.7% AutoSeq Frag,... - Mon, Sep 27, 2004 4:10:05 PM

All the sequences and contigs added to this particular sequencing project are displayed in this window.

Note: If you have so many open windows on the screen that it's difficult to find the Project Window, go to the **Window** menu and click on **Project Window**. This will bring the current project to the front.

As you work, if you close an existing project and then want to start a new project, go to the **File** menu and choose **New Project**. Sequencher will open a new untitled project.

Sequencher will remember the last projects you opened. You can see these as a list by going to the **File** menu and selecting the **Open Recent** command.

Opening an existing project

To open an existing project, go to the **File** menu and choose **Open Project** to select the project you want to work with. Click the **Open** button in the lower right area of the dialog box, or click the **Return** key.

Working with items in the Project Window

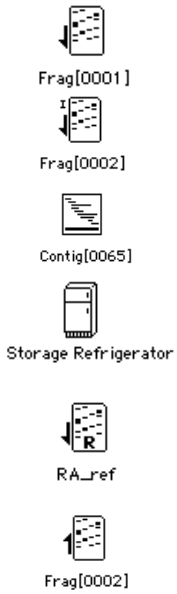
To work with items in the Project Window, first select the item(s) you need by clicking on them. Then choose the appropriate menu command.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Icon types

In the **Project Window** a sequence is represented by an autorad icon. However, there are other data types you may have in your project. To get familiar with some of the icons Sequencher uses for these data types, see Figure 3-3.

Figure 3-3 Examples of Sequencher icons



A sequence fragment

A sequence fragment with edited comments

A contig

A refrigerator

A Reference Sequence

A sequence in the inverse-complement orientation

Other information may be displayed in addition to the basic icon. For example, the downward-pointing arrow on a sequencing image icon shows that this sequence is stored in its original orientation (the sequence travels "down" the gel). The arrow points up when the data is inverse-complemented as shown in Figure 3-3 above.

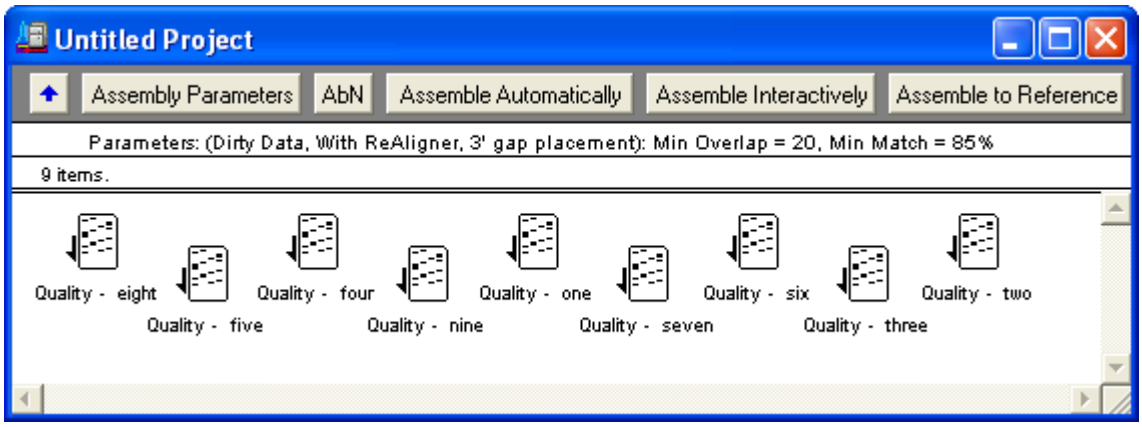
Project views

View as Large Icons

To display large icons in your Project Window, go to the **View** menu and then choose **Project Icons As.** Then select **Large Icons** from the submenu. The Project Window then displays the sequences and contigs as shown in Figure 3-4.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Figure 3-4 Project Window with Large Icons

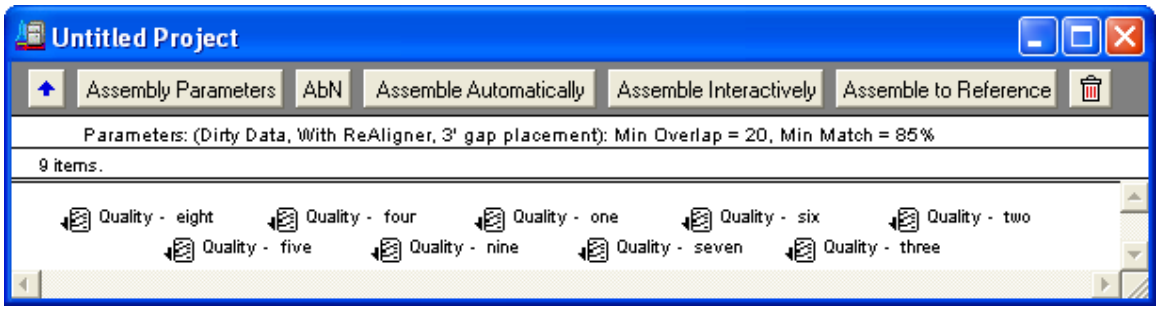


Note: A new project defaults to the icon view unless you previously specified a different view.

View as Small Icons

To fit more information onto the screen, you can view project items as smaller icons. Go to the **View** menu and choose **Project Icons As** and then choose **Small Icons** from the submenu. The Project Window then displays the sequences and contigs as shown in Figure 3-5.

Figure 3-5 Project Window with Small Icons



Moving icons

To reposition a sequence or contig icon, click the icon and drag it to the new location.

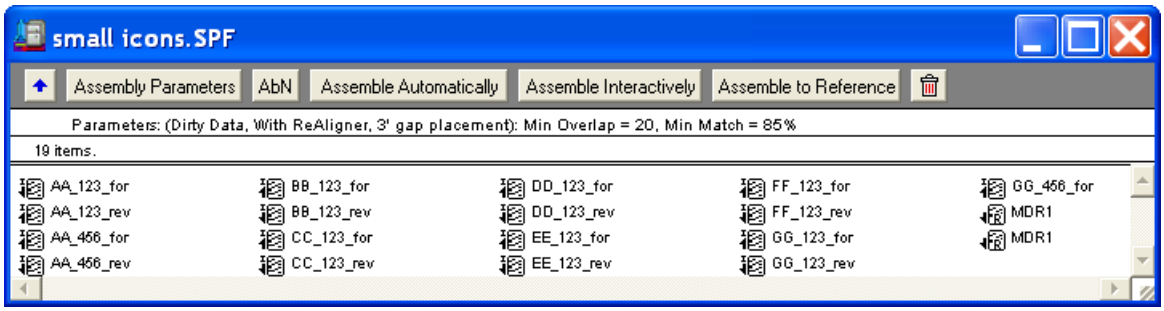
Note: Icons can be positioned on top of each other, thereby obscuring one or more of them from view.

Cleaning up

Sequencher has clean-up commands to help you organize the icons in your Project Window. Go to the **View** menu and choose the **Sort/Cleanup** command to see a submenu

that will let you rearrange the icons according to date edited, name, kind, size, etc. (See Figure 3-6).

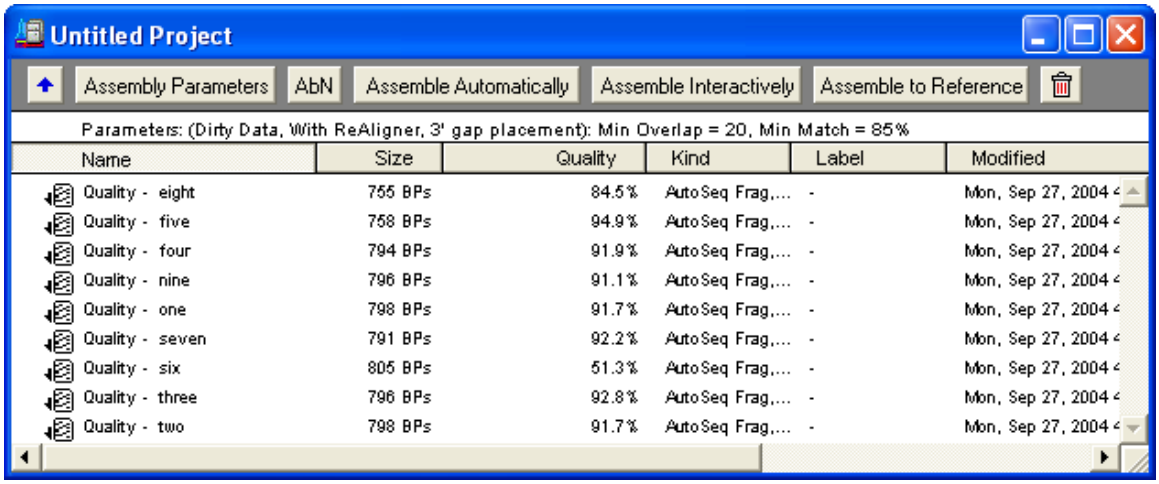
Figure 3-6 Small icons sorted by name



View as a list

You can show your project items as a list rather than as icons. Go to the **View** menu and choose **Project Icons As**… and then select **A List** from the submenu. The Project Window will look like the one shown in Figure 3-7.

Figure 3-7 Project Window with list view


Untitled Project
Assembly Parameters
AbN
Assemble Automatically
Assemble Interactively
Assemble to Reference
Parameters: (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20, Min Match = 85%
Name
Size
Quality
Kind
Label
Modified
Quality - eight 755 BPs 84.5% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - five 758 BPs 94.9% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - four 794 BPs 91.9% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - nine 796 BPs 91.1% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - one 798 BPs 91.7% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - seven 791 BPs 92.2% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - six 805 BPs 51.3% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - three 796 BPs 92.8% AutoSeq Frag,... - Mon, Sep 27, 2004
Quality - two 798 BPs 91.7% AutoSeq Frag,... - Mon, Sep 27, 2004

You can navigate the list by using the up and down arrow keys. To select a continuous list of sequences, choose the first sequence in the list and then, holding down the **Shift** key, select the last sequence in the list. To select a discontinuous list of sequences, hold down the **Control** key while clicking on the sequences.

Project Window Columns

When you are in the list view, Sequencher displays the attributes, such as the item name or kind, of each sequence as a series of columns. You can control which columns are displayed. Go to the **View** menu and choose the **Project Window Columns** command. Then choose your desired option from the submenu.

If you are working with ABI sequences, you can view the sample name by going to the **Project Window Columns** command and choosing the **Sample** command.

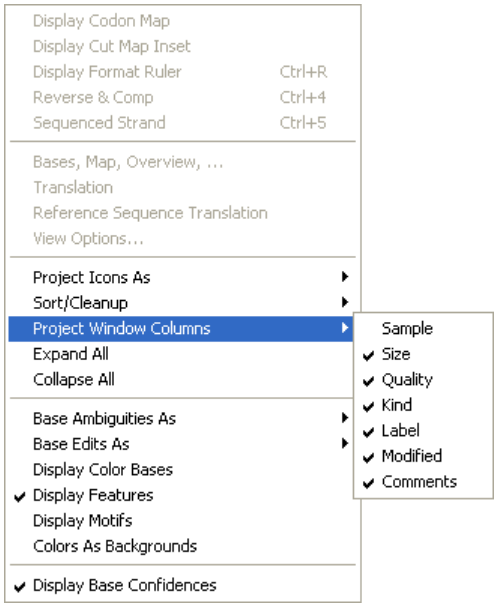
SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

The column labeled **Quality** displays a value for each sequence that has confidence scores. The Quality value is the percentage of bases in a sequence that is above the **Low** Confidence Range threshold. The default setting is 20. You can alter the **Low** threshold value in the **Confidence** pane of the **General** settings. See Chapter 21 "Customizing Sequencher and User Preferences" for more details.

You can sort your list by item name, size, quality, kind, label or modification date. Click on the title of the column you want to use for your sort. For example, to sort by size, click the word **Size** at the top of that column. Sequencher will reorder the list by size. You can also sort the list by going to the **View** menu, choosing **Sort/Cleanup** and then selecting from the options on the submenu.

If you sort by **Kind**, the order of precedence is first an unassembled Reference Sequence, or any Contigs which contain Reference Sequence fragments, then Contigs without Reference Sequences and finally Sequence Fragments at the bottom of the list.

Figure 3-8 Project Window Columns submenu



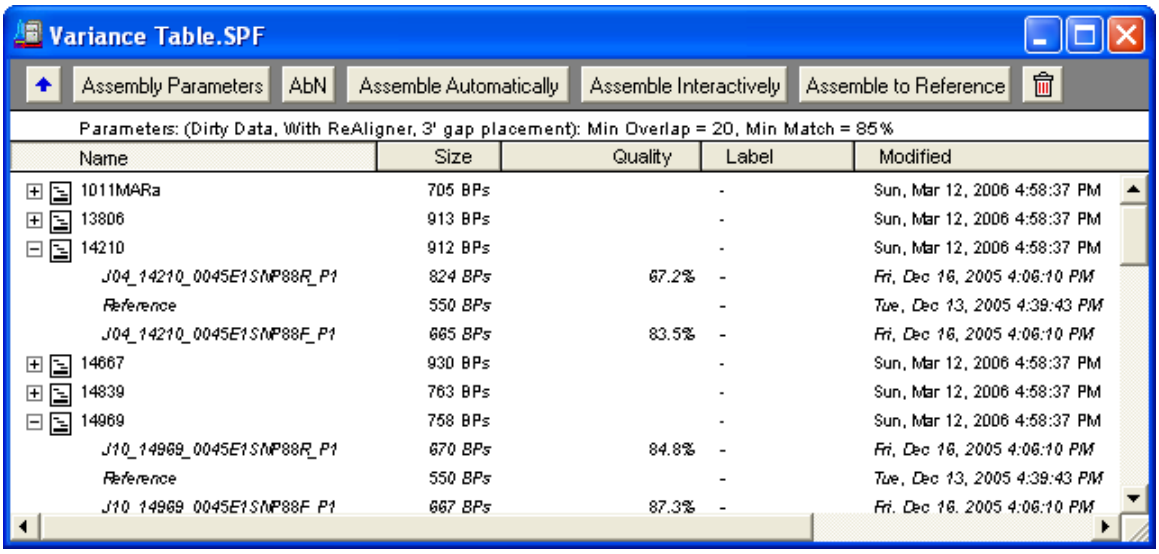
Note: You can toggle between types of views if you hold down the **Alt** key and click in the header area, just above the icons and below the parameter information.

Viewing constituent sequences

When you have contigs in list format, click the plus sign (+) to the left of the contig icon (Figure 3-9) to see a sublist of the sequences incorporated into that contig. That plus sign will turn in to a dash sign (-) and the sequence list will expand (see Figure 3-9). To hide a sublist of sequences, click the dash sign (-) again; it turns to a plus (+) and the sequence list will collapse.

Go to the **View** menu and click on the **Expand All** and **Collapse All** commands to either display or hide all of the contents of the project's contigs.

Figure 3-9 Viewing a sublist of a contig's constituents

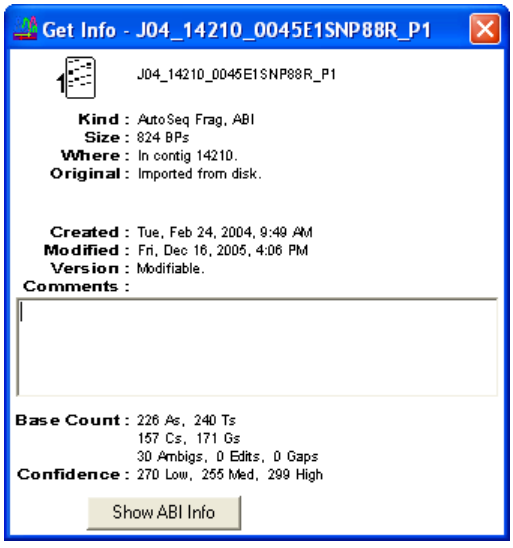

Variance Table.SPF
Assembly Parameters
AbN
Assemble Automatically
Assemble Interactively
Assemble to Reference
Parameters: (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20, Min Match = 85%
Name
Size
Quality
Label
Modified
1011MARa 705 BPs - Sun, Mar 12, 2006 4:58:37 PM
13806 913 BPs - Sun, Mar 12, 2006 4:58:37 PM
14210 912 BPs - Sun, Mar 12, 2006 4:58:37 PM
J04_14210_0045E1SNP88R_P1 824 BPs 67.2% - Fri, Dec 16, 2005 4:06:10 PM
Reference 550 BPs - Tue, Dec 13, 2005 4:39:43 PM
J04_14210_0045E1SNP88F_P1 665 BPs 83.5% - Fri, Dec 16, 2005 4:06:10 PM
14667 930 BPs - Sun, Mar 12, 2006 4:58:37 PM
14839 763 BPs - Sun, Mar 12, 2006 4:58:37 PM
14969 758 BPs - Sun, Mar 12, 2006 4:58:37 PM
J10_14969_0045E1SNP88R_P1 670 BPs 84.8% - Fri, Dec 16, 2005 4:06:10 PM
Reference 550 BPs - Tue, Dec 13, 2005 4:39:43 PM
J10_14969_0045E1SNP88F_P1 667 BPs 87.3% - Fri, Dec 16, 2005 4:06:10 PM

Viewing the information for a sequence or contig

The **Get Info** command displays additional information about a selected or open sequence or contig.

To view additional information about an open sequence or a contig editing window, go to the **File** menu and choose **Get Info**. If you do not have an open sequence or contig editing window you can select the icons of your sequences or contigs from the Project Window and then go to the **File** menu and choose **Get Info.**

Figure 3-10 The Get Info window


Get Info - J04_14210_0045E1SNP88R_P1
J04_14210_0045E1SNP88R_P1
Kind : AutoSeq Frag, ABI
Size : 824 BPs
Where : In contig 14210.
Original : Imported from disk.
Created : Tue, Feb 24, 2004, 9:49 AM
Modified : Fri, Dec 16, 2005, 4:06 PM
Version : Modifiable.
Comments :
Base Count : 226 As, 240 Ts
157 Cs, 171 Gs
30 Ambigs, 0 Edits, 0 Gaps
Confidence : 270 Low, 255 Med, 299 High
Show ABI Info

If you have sequences from an ABI sequencing system, Sequencher can display information from the ABI data sheet. In the **Get Info** window shown in Figure 3-10, note the **Show ABI Info** button.

Editing the information for a sequence or contig

You can annotate the sequence by clicking anywhere in the **Comments:** box and typing the information. To store more information than the **Comments:** box allows, go to the **Edit** menu and choose **Edit Comments…**. Sequencher opens a small text editing window that stores about 250 characters of text for each sequence. After you have added text with **Edits Comments…** a small "I" (for "Info") appears in the upper left corner of the icon. If you want to remove the information in the **Comments:** box, select the text and then go to the **Edit** menu and choose **Contents** from the **Clear** option.

The Project Window button bar

The button bar contains buttons for frequently-used commands. It appears just below the top of the Project Window. You can use the smaller left-hand button, which has an up arrow on it, to hide the button bar. Click on the button and the button bar will "hide" behind the title bar.

When the button bar is hidden, you can still see the bottom of the left-hand button under the title bar. It will show a down-pointing arrowhead. Click on the arrowhead and the button bar will reappear.

If you click on the **Assembly Parameters** button this will display the **Assembly Parameters** window. Clicking on the **AbN** button will enable the **Assemble by Name** function. **Assemble Automatically**, **Assemble Interactively** and **Assemble to Reference** are used to assemble sequences you have selected. If **AbN** is enabled then the **Assemble Automatically** and **Assemble to Reference** buttons will have different labels. If you select a sequence and then click on the **Trash Can** button, that sequence will be removed from the project.

Templates

A template is a special type of project that can be set with your choice of parameters and preferences. Templates can contain ordinary sequences or a Reference Sequence. Preferences can include Feature settings, enzyme sets, and display settings. With templates, you can also set **Assembly Parameters**, **Assemble by Name** handles and **Trim** parameters. These will be discussed later in this manual.

A template can be used as a new project or imported into an existing project. You can use templates to set up standard operating procedures and methods to reuse or distribute throughout your lab.

Save Project As Template

You can save any project as a template by using the **Save Project As Template** command. Set the **User Preferences** and parameters that you want to use. You can include a Reference Sequence or ordinary sequences.

Once you are satisfied with your choices go to the **File** menu and click on the **Save Project As Template…** command. A new dialog window called **Template Name:** will appear. Type a name for your template in the text box and then click on the **OK** button. Your project will now be saved in a Templates folder. The name of the template will appear in the **New Project From Template** submenu.

New Project From Template

The **New Project From Template** command allows you to open a new project containing all the sequences, settings and preferences associated with your chosen template. Go to the **File** menu and click on **New Project From Template**. Select a template from the submenu. A new blank Project Window will open.

Import from Template

You can add preferences, settings and sequences to an existing project by importing a template. If you have an open project, go to the **File** menu and click on the **Import command** and then choose **From Template**. Select a template from the submenu. The settings and/or sequences from the template will be applied to your open project automatically.

4. Importing data

In this chapter we will discuss how to import data from automated sequencing equipment and other programs into Sequencher projects. The different techniques we will discuss include dragging and dropping, copy and paste and using import commands. We also explain how to deal with projects in specific formats, how to import confidence scores and how to remove unwanted sequences and contigs from a project.

Quick and easy importing of data

Dragging and dropping files

You can drag files from the desktop right into the **Project Window** and Sequencher will import and load those files. Sequencher will alert you if it cannot read and load the files.

If you drag sequences onto the Sequencher icon on your desktop you will open a new **Project Window** containing your sequences. If Sequencher is already open, the sequences will be imported into the open project.

***Note*:** Sequencher always shows a newly imported file as selected, so if you do not want to use it right away, de-select it by clicking elsewhere in the project.

Copy and paste

One of the easiest ways to move data between Sequencher and other programs is to use the clipboard's **Copy** and **Paste** commands. First, launch both Sequencher and the other program. Then, select and copy the bases of interest in the other program.

Go to the **Sequence** menu and choose **Create New Sequence**. Type a name for your sequence into the **Sequence's name:** box and then click the **OK** button. Now go to the **Edit** menu and choose **Paste**. When you close the window you will see a dialog box with two buttons. If you have finished entering your data, click the button **Record As Experimental Data**. Otherwise click the button **Not Yet Finished**. Sequencher then removes the window.

How to import using menu commands

Importing sequences

Sequencher imports sequences in a variety of text and chromatogram formats. Some of the most popular are ABI, MegaBase, CEQ, SCF, GenBank, EMBL, DDBJ and plain text files.

To import a sequence, first bring the Project Window to the front by clicking on it. Under **File** menu, choose **Import.** Choose **Sequences**… and then browse to and select the file you want to import. Click the **Open** button in the lower right area of the dialog box or press the **Return** key. You may need to select a file format from the drop down menu called **Files of Type:** in order to see the files of interest in the dialog window. To see all the files in a particular location choose **All Files (*.*)**.

You can also select from a **Folder of Sequences**…, **ACE Project**…, **CAF Project**…, **GCG contig**…, **Sequencher project**… or **Sequence from VecBase**…. (Working with specific formats is discussed below.)

If Sequencher cannot import the file you will get a warning message to that effect.

Importing GenBank features

When you import a sequence in GenBank format, you will automatically import the features, including the location, the feature keys and feature qualifiers. The feature key used in sequences published in GenBank, EMBL and DDBJ describes the biological nature of the annotated feature or indicates information about changes to the sequence. (For more information on features and feature keys see Chapter 17 “Motifs and Features” and Appendix 28 “Feature Keys and Qualifiers”.)

Where feature locations described in the GenBank file are ambiguous, Sequencher will warn you that it cannot import these features and will list the features in a dialog box you can print out for your records.

Importing lists of sequences

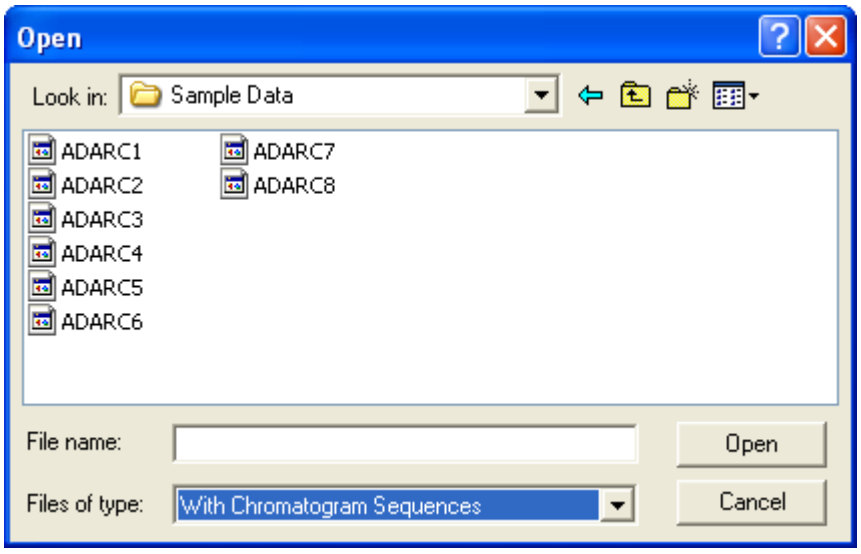
Sequencher’s functionality has been modified to take advantage of the MacOS X file system. If you use the **Import** menu item with the suboption **Sequences…** OSX allows you to select multiple files for a single import command. To select a continuous list of sequences, choose the first sequence in the list and then, holding down the **Shift** key, select the last sequence in the list. The import window selects the two sequences and all of the files in between. Click the **Open** button to import all of the selected sequences in to your project.

To select a discontinuous list of sequences, hold down the **Control** key while clicking on the file. Click the **Open** button to import all of the selected sequences into your project.

The **Files of Type:** drop down menu allows you to filter the files that are displayed in the file browser. The filters are **With Chromatogram Sequences, Without Chromatogram**

Sequences, Supported File Formats and **All Files (*.*).** You may see some files becoming grayed-out with different file filters because these files do not meet the import criterion of the current filter (see Figure 4-1where **With Chromatogram Sequences** is set but some of the files are text only files).

Figure 4-1 File import Enable option set to With Chromatogram Sequences



Note: To import traces you must import the actual trace file. If you import a text file containing ASCII characters, you will not import traces. Text files will be only a couple of kilobytes in size, whereas trace files will be well over 50 kilobytes.

Sequencher uses a very efficient algorithm to compress the imported traces. This conserves disk space so a project file containing several sequences from an automated sequencer may be smaller than any one of the original trace files.

Folder of sequences

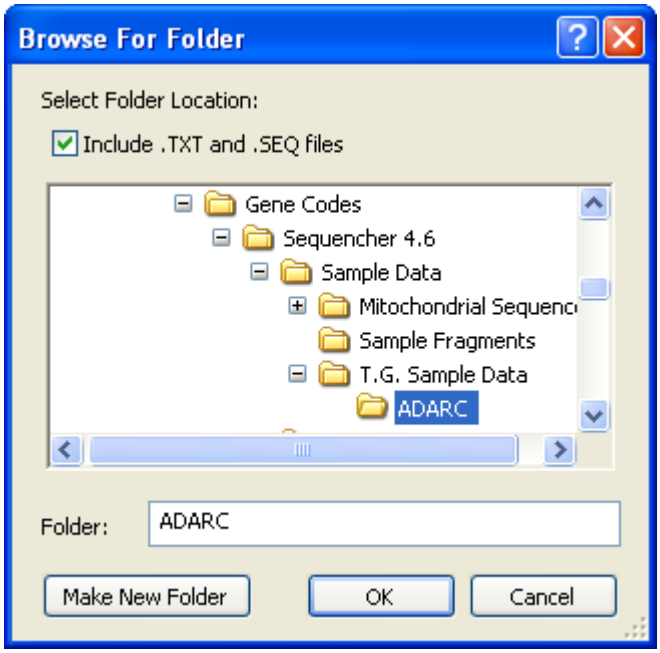
When you have many sequences to import you can import all those files from within a single folder. Select the Project Window and go to the **File** menu and choose **Import** and **Folder Of Sequences…** from the submenu. The standard **Browse for Folder** dialog box appears; select the folder that holds the data you want to import and click **OK**.

Use the **Include .TXT and .SEQ files** check box to include or exclude these files from your import.

Sequencher tells you how many files are in the folder and asks whether you want to import them all. If you do, click **Import All Files in Folder**. Sequencher imports all the files, assigning a separate icon to each.

You can specify that you want files imported into the Project Window or directly into the trim window (see Chapter 6 “Preparing your Data for Assembly”).

Figure 4-2 Browse for Folder dialog box



Projects in specific formats

Sequencher supports importing projects as ACE projects, GCG contigs and as CAF projects. The contents of these files include confidence values, feature annotation and assembly information. Select the Project Window and go to the **File** menu and choose **Import** and then choose the specific project format from the submenu. The standard **Open** dialog box appears; select the folder that holds the data you want to import and click **Choose.**

Read more about importing confidence values in the section "Importing Confidence Scores" later in this chapter.

Sequencher projects

At times you may want to merge two projects or combine your own work with that of a colleague. To combine Sequencher projects, first open one of the projects. Go to the **File** menu and choose **Import** and then **Sequencher Project…** from the submenu. Select your other project and click the **Open** button. Sequencher displays all the imported projects in the initial Project Window, retaining all comments, edits and even the relative positions of the icons. To keep track of which sequences came from which project, you may want to establish a naming convention for sequences *before* combining them. Since the imported sequences and contigs are highlighted after import you could also use sequence **Labels** to differentiate them. (See Chapter 21 "Customizing Sequencher and User Preferences".)

Files from VecBase

VecBase is a file of vector sequences in your Sequencher application folder. You can import a vector from VecBase just as if it were any other DNA sequence. Go to the **File**

menu and choose **Import** then the subcommand **Sequence From VecBase**. The **Vector Selection** dialog box appears; choose the desired file and click **Select**. If you wish to import more than one file you can use **Shift** or **Command** click for continuous or discontinuous listings of sequences.

Creating a new sequence

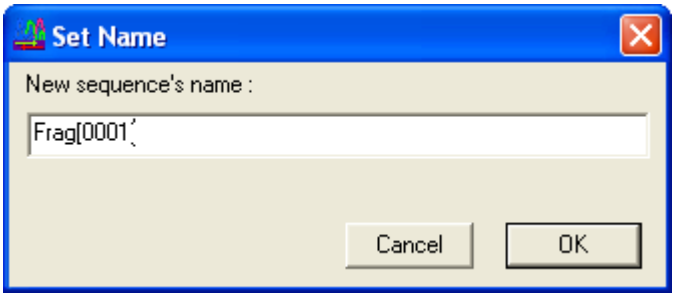
Another way to bring new data into Sequencher is to create a new sequence fragment. To do this, go to the **Sequence** menu and choose **Create New Sequence…**

Sequencher displays a dialog box asking you to name the new sequence (Figure 4-3). The default name, in this case "Frag[0001]," can be set as a user preference. A sequence name can be a maximum of 255 characters, including any punctuation and spaces, but 31 characters is recommended for export functions.

When you have typed the name you want (or if you accept the default name), click **OK** or press the **Enter** key. A new Sequence Editor appears showing the name you selected in the title bar. The new Sequence Editor lets you enter and change sequence data. Just type as you would with a standard word processor.

When you are finished entering and/or editing a sequence, click the close button in the top left corner of the window at the left end of the title bar or go to the **File** menu and choose **Close Window**. Sequencher asks if you have finished editing. At this point you can record your data as experimental data or you can tell Sequencher you have not yet finished editing by selecting the appropriate button. Sequencher then removes the Sequence Editor window from the screen. The sequence is now represented by a sequencing icon in the Project Window.

Figure 4-3 Create New Sequence dialog box



Double-clicking a project icon

When you double-click on a sequence project icon, this launches Sequencher and opens the project. If a Project Window is *already* open, when you double-click on a project icon Sequencher will ask whether you want to save changes to your current project before closing it. Click the **Cancel** button if you do not want to proceed. If you want to close the project and save any changes you have made, click the **Yes** button. If you want close the project without saving your changes, click the **No** button.

Importing Confidence Scores

Many base callers generate confidence values associated with each base call. Sequencher supports confidence values from PHRED, ACE and Trace Tuner files, as well as SCF, ABI 3730 and ESD files.

Importing PHRED called data

Sequencher can read in the SCF, FASTA and Qual files generated by PHRED and supports PHRED base calls and confidence values. To view the confidence levels, you must first set the user preferences for displaying confidence values (See Chapter 21 "Customizing Sequencher and User Preferences" for more information). Then go to the **View** menu and choose **Display Base Confidences**.

To import the data into Sequencher, it must first be organized in a folder containing three subfolders. These subfolders are chromat_dir which contains the trace files, edit_dir which contains the text files and phd_dir which contains the quality scores. Select **File** then go to **Import** and then **Folder of Sequences** to import the phd_dir folder. To import the sequences you would like to view you may also go to the **File** menu and choose the **Import** command with the **Sequences...** submenu.

Importing PHRAP Files

Sequencher can import the ACE projects created by PHRAP. To import an ACE project select **File** then go to **Import** and then the **ACE Project…** submenu. Browse to a file that ends with ".ace" or ".ace.x" (such as .ace.1,.ace.2, etc.). This file will typically be located in the subfolder "edit_dir." Select the highest numbered ACE file to account for all editing and reassembling.

Sequencher automatically imports and incorporates as much auxiliary data as it can find. The auxiliary files must be in the same folder as the ACE file or arranged in the traditional PHRAP assembly hierarchy (chromatogram files in a chromat_dir folder and PHD files in a phd_dir folder). If the .singlets file, the .qual file, the PHD files and the chromatogram files all exist and are available, Sequencher will import them.

Alternatively, you can import parts of the PHRAP project by going to the **File** menu and choosing **Import** and then **Sequences** or by going to the **File** menu and clicking on **Import** and then **Folder of Sequences**.

- If you import the chromatogram files from the "chromat_dir", these fragments will have PHRED base calls and chromatogram data but no base qualities.
- If the "_.phd" files are loaded and "_. fasta.qual" files are available, you will have imported all of the fragments from the PHRAP project. They will have base quality and chromatogram data.
- The.qual files alone will not load your sequences.
- If you import PHD files from the "phd_dir" and the corresponding chromatogram files are available, these files will import with PHRED base calls, chromatogram data and base qualities.

Note: When using ftp to transfer your data to your local computer, it is ALWAYS important to transfer these files in *binary* mode. If you are not familiar with ftp, speak with your network administrator.

Removing sequences and contigs from a project

To remove items from the project, select the items you want to remove. Go to the **Edit** menu and choose **Remove From Project…**, or click the **Trash Can** button in the button bar. Removing an item from a project is permanent, so Sequencher asks whether you really want to do it.

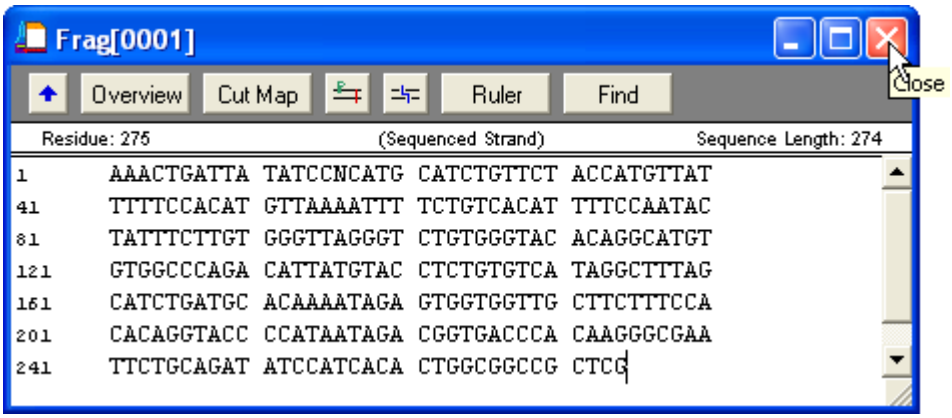
***Note*:** Remember, you cannot back out of this action with a simple **Undo** command, but if you close the project *without saving,* all changes including removed items will be forgotten. The **Revert to Saved Project** command has the same effect.

Closing an editor window

When you have finished entering and/or editing a sequence, click the close button in the top right corner of the window at the right end of the title bar (see Figure 4-4) or choose **Close Window** from the **File** menu. Sequencher then removes the Sequence Editor window from the screen.

Once you have closed the window, a sequencing icon will represent your sequence. (Figure 4-4).

Figure 4-4 Clicking the close box



Closing a project

If you have finished working with the project but not with Sequencher, choose **Close Project** from the **File** menu. Sequencher will ask if you want to save the changes to your project. Click **Yes** to save your changes and close the project. Click **No** to close the project without saving your changes. Click **Cancel** if you wish to continue working with the project. You can then start working with another project by going to the **File** menu and choosing **New Project** or **Open Project**.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

If you have finished working with Sequencher, go to the **Sequencher** menu and choose **Quit Sequencher**.

5. Organizing project views

In this chapter on organizing project views you will learn more about the structure of a Sequencher project as you begin to work with your data. We will discuss how to select individual and multiple sequences using various techniques, renaming items and saving your work.

Working with icons and lists

Changing between project views

Sequencher provides you with a variety of ways to view and organize the contigs and unincorporated data in your Project Window. You can view your data as large or small icons or in list form (for more information see Chapter 3 "The Project Window"). You can toggle quickly between types of views by holding down the **Alt** key and clicking in the header title area, just above the icons and below the parameter information. You can display a list of sequences in a single contig in the list view by clicking on the plus sign to left of the contig icon. You can see a list of sequences in all your contigs by going to the **View** menu and choosing **Expand All**.

Collecting sequences in refrigerators

Sequencher lets you create "refrigerators" in the Project Window to hold certain subsets of your project. Refrigerators hold only raw sequences, not contigs or other refrigerators. To "put away" a few sequences in "cold storage," select the sequences and then go to the **Edit** menu and choose **Refrigerate.** Sequencher collects the sequences and puts them in the refrigerator.

To remove sequences from the refrigerator, open the refrigerator by double-clicking on it. You will see a list of items. Select the items you want to remove and then click the button called **Move Selected Items To Project Window** to execute the command.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Selecting sequences

Selecting sequences and contigs

To select sequences you can go to the **Select** menu and use the **Select All, Select None** and **Invert Selection** commands. **Invert Selection** will give you the opposite of your previous selection.

There are also selection functions that are increasingly specific. You can choose data based on **All Items That…** and then further specify the data by choosing narrower classes such as **Contain Subsequence…, Contain Items Named…** or **Have Chromatograms** from the **Select** menu. You can indicate if you want these names to be case sensitive or not.

Note The subcommand **Contain Items with Names Containing**… only works on contigs.

For more detailed information on these commands, see Chapter 18 “Finding Items”.

Selecting with the mouse

First make sure the Project Window is at the front. When you click on an icon, or hold down the **Shift** key while clicking on several icons, Sequencher highlights the icon(s). When you have finished selecting a group of icons, release the **Shift** key before you try to do anything with the group.

You can deselect an icon by clicking on it while holding down the **Ctrl** key. If you do not hold down the **Shift Ctrl** key, Sequencher deselects the last highlighted icon(s) as soon as you click on another. If you click in the white space *between* icons and names, Sequencher deselects all the highlighted icons.

When you display the list view of a Project Window or a contig you can select a range of fragments by using **Shift click.** You can make a discontinuous selection by using the **Ctrl** key and clicking on the individual fragments.

Selecting by typing

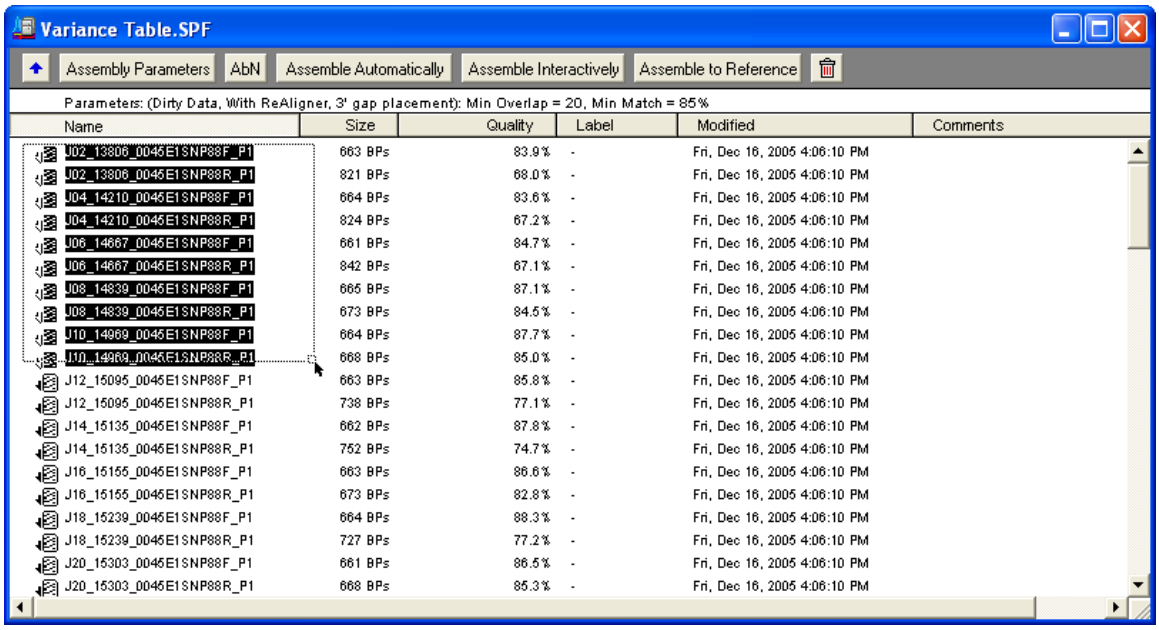
Type the name of a sequence fragment or contig. As you type, Sequencher searches the Project Window for an icon with that name. If it finds the name, it scrolls to the point at which the icon is visible and highlights the icon.

Note: If you pause for more than a second or two while typing a name, Sequencher assumes you have started typing a different name and starts looking for the new one instead.

Selecting multiple items with a marquee

You can select multiple sequences by drawing a box around the icons or list items you want. Position the mouse so it is not touching any icons or names. The cursor should be approximately where one corner of the box should be. Hold the mouse button down and drag diagonally across to where you would expect the opposite corner of the box. As you drag, the marquee appears (as shown in Figure 5-1). As you move the mouse, any icons or names touched by the marquee will be selected

Figure 5-1 Selecting with the marquee



Selecting using menu commands

Selecting all icons

You can change the icons selected in the Project Window by going to the **Select** menu and clicking on the **Select All, Select None** and the **Invert Selection** commands.

Selecting all items containing sequence

To select all items in the Project Window that contain a particular subsequence, go to the **Select** menu and choose **All Items That** command. Then select **Contain Subsequence…** from the submenu. Sequencher displays a dialog box that lets you type two sequences; the box has buttons to specify how you want the sequences matched to items in the Project Window. You can use the drop down menus to specify restriction enzyme recognition sequences.

Finding assembled sequences

To locate data assembled in a contig go to the **Select** menu and click on one of the following options from **All Items That** then **Contain Subsequence…**, or **Contain Items With Names Containing…**. Sequencher will select the contig(s) that match your selection criteria.

You can also use the **Select** and then **Item Named…** menu option to locate which contig contains a particular fragment. With this command you must type the full name in to the **Find What:** box and then press the **Find** button. The contig containing the fragment will be highlighted.

Renaming a sequence or contig

Renaming using the mouse

To rename a sequence fragment or contig with the mouse, click twice on the name of the item. This selects the icon and puts a text box around the name. Just type in the new name and when you are done, click elsewhere in the Project Window.

If you only want to change part of the name, drag the cursor over what you want to replace and then type the replacement text in the highlighted area.

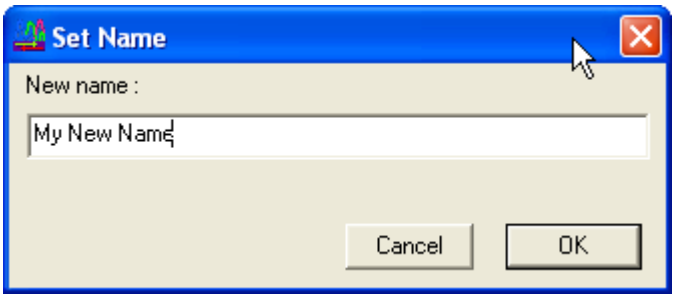
You can also use the **Copy** and **Paste** functions to rename sequences.

Renaming using the menu

If you wish to rename a sequence you can go to the **Sequence** menu and choose the **Rename Sequence…** command. To rename a contig, go to the **Contig** menu and click on the **Rename Contig…** command. Sequencher displays a dialog box (Figure 5-2) with a field for you to enter the new name.

Click the **OK** button when you have finished.

Figure 5-2 Renaming a sequence



Labeling items in the Project Window

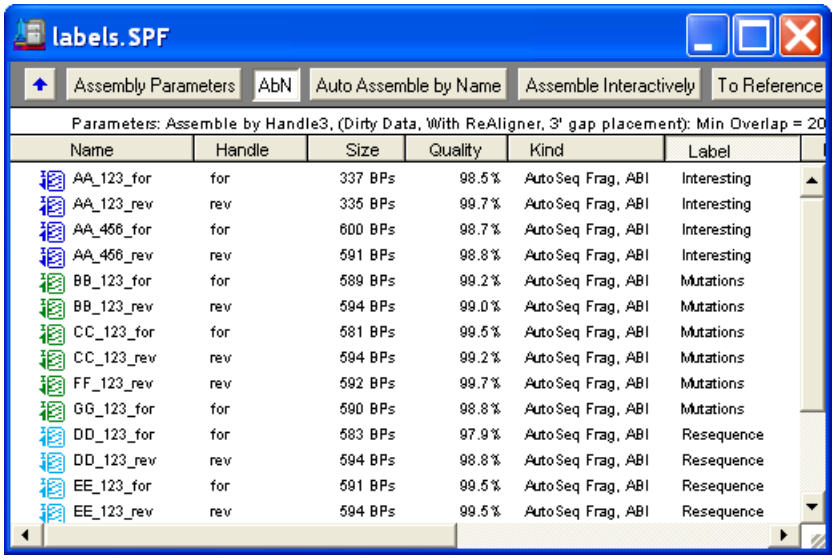
You can organize items visually in the Project Window by applying **Labels** to the icons. Before using **Labels** for the first time, you should customize the list of labels under the **General** section of **User Preferences**. You will see a list of **Label & Name** options. You can also select from the default labels available under the **Edit** menu and **Label** submenu. (See Chapter 21 “Customizing Sequencher and User Preferences” for more information.)

When viewing your project as a list, you will see the name of the label in the list.

Labeling your icons allows you to sort or find your information easily. For instance, you might define red labels as cDNA data and blue labels as genomic data. Clicking on the label heading in the Project Window will sort your project items by label (see Figure 5-3).

You can also find sequences bearing a particular label by going to the **Select** menu and choosing the **All Items That** command with the **Have Labels Containing…** suboption.

Figure 5-3 Viewing labels as a list


labels.SPF
Assembly Parameters
AbN
Auto Assemble by Name
Assemble Interactively
To Reference
Parameters: Assemble by Handle3, (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20
Name
Handle
Size
Quality
Kind
Label
AA_123_for for 337 BPs 98.5% AutoSeq Frag, ABI Interesting
AA_123_rev rev 335 BPs 99.7% AutoSeq Frag, ABI Interesting
AA_456_for for 600 BPs 98.7% AutoSeq Frag, ABI Interesting
AA_456_rev rev 591 BPs 98.8% AutoSeq Frag, ABI Interesting
BB_123_for for 589 BPs 99.2% AutoSeq Frag, ABI Mutations
BB_123_rev rev 594 BPs 99.0% AutoSeq Frag, ABI Mutations
CC_123_for for 581 BPs 99.5% AutoSeq Frag, ABI Mutations
CC_123_rev rev 594 BPs 99.2% AutoSeq Frag, ABI Mutations
FF_123_rev rev 592 BPs 99.7% AutoSeq Frag, ABI Mutations
GG_123_for for 590 BPs 98.8% AutoSeq Frag, ABI Mutations
DD_123_for for 583 BPs 97.9% AutoSeq Frag, ABI Resequence
DD_123_rev rev 594 BPs 98.8% AutoSeq Frag, ABI Resequence
EE_123_for for 591 BPs 99.5% AutoSeq Frag, ABI Resequence
EE_123_rev rev 594 BPs 99.5% AutoSeq Frag, ABI Resequence

Saving your work

Saving your Project

To save a project go to the **File** menu and choose **Save Project**, which saves all your changes to the current project file. When you use the **Save Project** command for a new project, Sequencher will prompt you to name your project with the **Save As** dialog. Enter a name for the new project in the **File name:** field then click **Save.**

Auto-Save

If you are working on complex projects you may find it helpful to have Sequencher automatically save your work at pre-set time intervals. You will need to enable the **Auto-Save** user preference from **User Preferences…** in the **Window** menu. (See the Chapter 21 "Customizing Sequencher and User Preferences" for more details.)

Saving a duplicate of a current project

If you wish to create a second copy of a current project you can use the **Save Project As…** command to create the new copy of your project with a different name. When you save the new project, a new project icon appears on your disk.

Reverting to the saved version of the project.

If you are working in a project and want to cancel all the changes you have made since you last saved it, go to the **File** menu and choose **Revert To Saved Project…** Sequencher displays a dialog box that tells you when your project was last saved. If you click the **Yes, Revert to Saved Version** button you will lose the most recent changes. Click on the **Cancel** button to cancel the command.

Note: Remember, if you click the **Yes, Revert to Saved Version** button you cannot undo this action.

Closing the Project Window

When you have finished working on an open project and want to work on another, you must first close the open project. You can either close the project by going to the **File** menu and clicking on **Close Project** or you can close the Project Window by clicking on the close box. In either case, Sequencher will prompt you to save your work if you have made any changes since your last save.

To continue working in Sequencher, either go to the **File** menu to create a new project by choosing **New Project** or click on **Open Project** to work on a previously saved file.

Note: If you want to save the data you have entered, you must save the project. Saving a project is different from recording a sequence as experimental data. Only saving the project actually stores data on the disk!

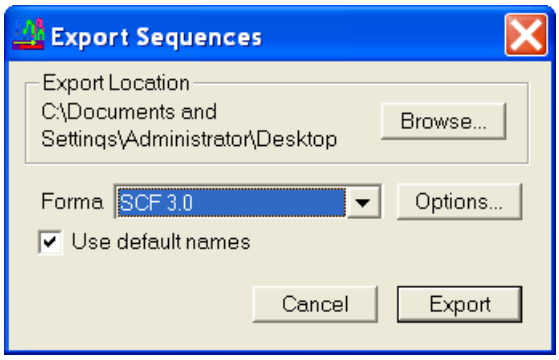
Exporting data

You can save your work in a file format other than Sequencher. You can export the entire file or a subset of the project. Click on the **File** menu and choose an **Export** submenu option. A dialog box will allow you to specify the file format in which you want to export your selection (Figure 5-4). For example, if you choose **Selection as Subproject** you can specify CAF format (for which you can set a number of options), Fasta concatenated and older versions of Sequencher. This is especially helpful if you want to share data with colleagues working with older versions of Sequencher.

Note: It is important to remember that any version-specific features you may have used will be lost if you save to an earlier version of Sequencher.

(See Chapter 19 “Exporting Data ” for more information.)

Figure 5-4 Export sequences format menu



Exiting the program

When you have finished working with Sequencher make sure you have saved your project. Then go to the **Sequencher** menu and choose **Quit Sequencher**.

6. Preparing Your Data for Assembly

In this chapter we introduce you to Sequencher's powerful tools for trimming poor quality data or vector contamination from your sequences. You will learn how to use these functions when you add sequence data to a project. We discuss the various criteria you can use such as trimming off ambiguities, trimming off low confidence data, how to screen for specific vector contamination and how to change all the trim tools' criteria to your own specifications.

Removing undesirable data

Sequencher's trimming tools allow you to trim one sequence at a time or thousands of sequences at a time. The two trimming commands under the **Sequence** menu are **Trim Ends…** and **Trim Vector...**. Both the **Trim** dialogs share a similar interface. All the sequences to be examined are collected in a single table. The criteria that you set determine how much data you trim off.

The trimmed data are fully recoverable because Sequencher always stores two copies of every imported file, the original sequence and the edited version. In this way you can always use the **Revert to Experimental Data** command from the **Sequence** menu to recover the original untrimmed sequence.

Trimming poor quality data

Selecting and trimming fragment data

Select one or more fragment(s) in the Project Window by clicking on them. Then go to the **Sequence** menu and choose the **Trim Ends…** command. Note that if you select contigs the **Trim Ends…** function only affects unassembled fragments. All the sequences are collected in to a single table. At the top of the table is a button bar (Figure 6-1), that you use to manage trim criteria, review which bases will be trimmed and perform the trim command.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Performing a trim with default settings

You can specify how much dirty data should be trimmed off as a function of the number of ambiguous base calls per run of bases by altering the **Ends Trimming** criteria. Each pane in the **End Trim window** shows how much sequence should be trimmed off. There will be a checkbox at an end if it requires trimming. Bases to be trimmed will be marked in red.

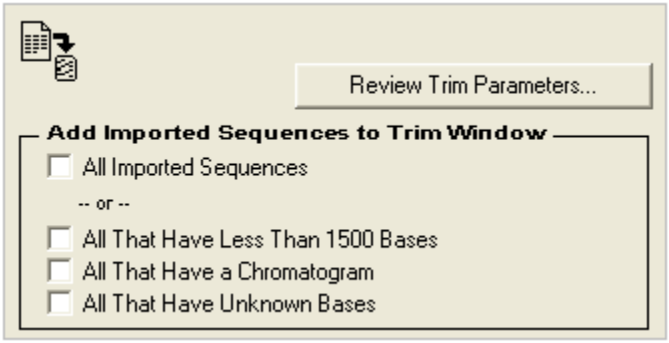
You can proceed with the trim by clicking on the **Trim Checked Items** button. Once you have trimmed your data dismiss the window with the command from the **File** menu called **Close Window.**

Figure 6-1 The Ends Trimming button bar

Automatically selecting sequences for ends trimming

You can set the **User Preferences** so that sequence files are automatically added into the trimming window as they are imported. During **Import**, Sequencher loads the sequences that require trimming directly into the trim window. To determine which sequences are directly imported into the trim window go to the **Window** menu and under **User Preferences** click on **Input/Output** and then select the submenu **File Import**.
Figure 6-2 shows the criteria you use to specify which sequences are candidates for direct import to the **Ends Trimming** window. (For more information see Chapter 21 "Customizing Sequencher and User Preferences".)

Figure 6-2 User Preferences for setting trim criteria when files are imported

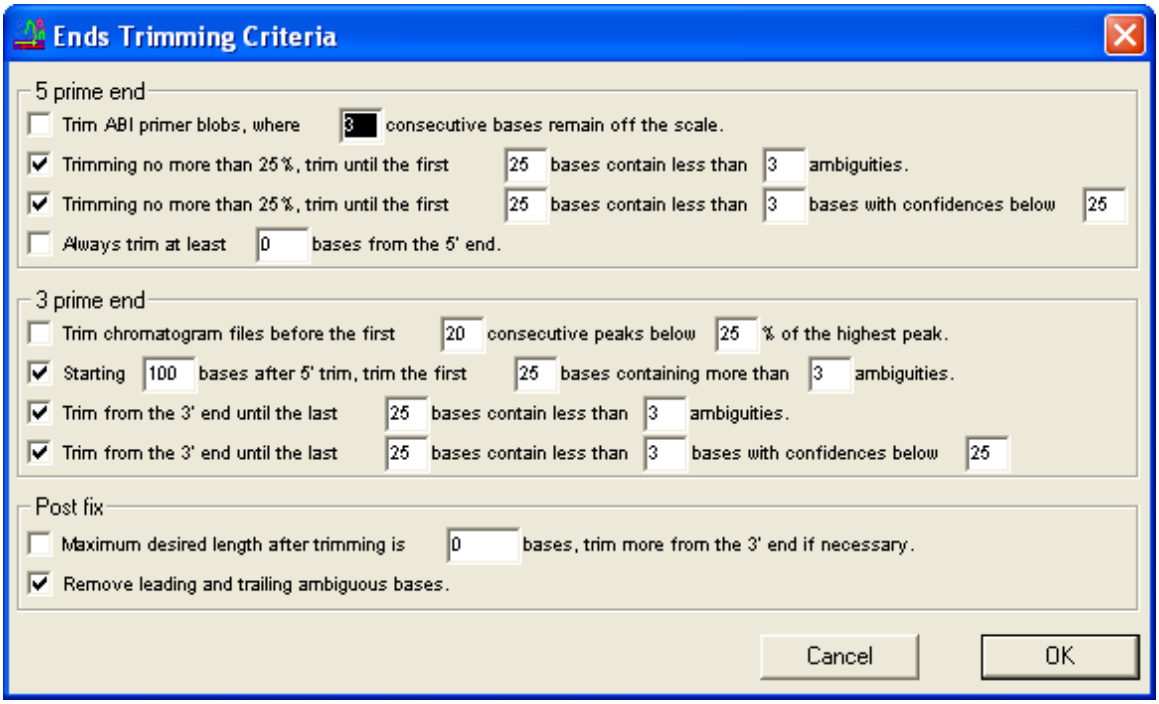

Review Trim Parameters...
Add Imported Sequences to Trim Window
All Imported Sequences
-- or --
All That Have Less Than 1500 Bases
All That Have a Chromatogram
All That Have Unknown Bases

How to set the Ends Trimming criteria

The amount to be trimmed is based on the criteria you set for both the 5' and 3' ends of the sequence. These criteria can be set independently of each other. To review or change those criteria, go to the **Ends Trimming** window and click on the **Change Trim Criteria** button. As you can see in Figure 6-3, you can trim off dye-primer peaks, set windows of acceptable reliability, percentage of the highest peak height (3' end) and even set an absolute number of bases. For example, you could specify that fragments must have

fewer than 3 N's in the first 20 and last 30 bases. To activate specific trim criteria you must first select the checkbox that governs the criteria fields.

Figure 6-3 The Ends Trimming Criteria window


Ends Trimming Criteria
5 prime end
Trim ABI primer blobs, where 3 consecutive bases remain off the scale.
Trimming no more than 25%, trim until the first 25 bases contain less than 3 ambiguities.
Trimming no more than 25%, trim until the first 25 bases contain less than 3 bases with confidences below 25
Always trim at least 0 bases from the 5' end.
3 prime end
Trim chromatogram files before the first 20 consecutive peaks below 25 % of the highest peak.
Starting 100 bases after 5' trim, trim the first 25 bases containing more than 3 ambiguities.
Trim from the 3' end until the last 25 bases contain less than 3 ambiguities.
Trim from the 3' end until the last 25 bases contain less than 3 bases with confidences below 25
Post fix
Maximum desired length after trimming is 0 bases, trim more from the 3' end if necessary.
Remove leading and trailing ambiguous bases.
Cancel
OK

Using Confidence to Trim Ends

In addition to trimming data based on the number of ambiguities at the 5' and 3' ends, you can trim sequence data based on the confidence scores assigned by your base caller. Confidence scores (also called quality values) are numbers associated with each base call and which define the likelihood that a base call is incorrect. The most common scale is from 1-60, where "60" represents a $1/10^6$ chance of a wrong call and 20 represents a $1/10^2$ chance. Depending on the program used, the confidence score may be based on peak height, the presence of more than one peak, and/or the spacing between the peaks. (See Figure 6-3 the **Ends Trimming Criteria** window for trimming by confidence criteria.)

Reviewing the Trim

Each pane in the trimming windows shows schematically how much sequence should be trimmed off. Blue lines represent the acceptable sequence data. The blue line is then flanked by red scissors, which mark both the 5' and 3' suggested trims based on the current **Trim** criteria. The remaining red regions of the sequence define the portion of the sequence to be trimmed.

Click the **Show Bases** button to see more information on the bases. After you have clicked **Show Bases**, the button toggles to **Show Overview**. Sequencher also provides additional control within the **Trim Ends** window. If you want Sequencher to ignore the

trim command for one or the other ends of the fragment(s) without modifying the overall specifications, just click off the checkbox associated with that trim position.

Figure 6-4 shows a portion of the base detail for these sequences. You can double-click on the sequence icon and invoke a Sequence Editor to examine the data in more detail.

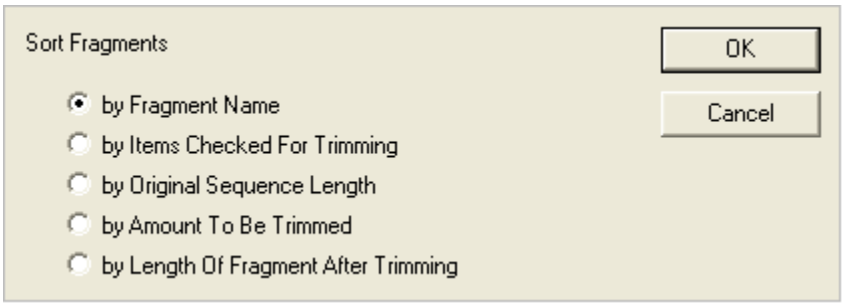
Figure 6-4 Bases view in Ends Trim window

Ends Trimming
Trim Checked Items
Change Trim Criteria
Sort Items
Show Overview
9 items in list
9 items have at least 1 check
Quality - eight (Has Chromatogram)
585 good bases, (2 Ambigs)
...nntggnnnncnnnnnaggancccntca CGGCCTGAGGAAGGGTGG...CGTCCTCACTAGCCTCCT cgctcacggtagaccaagaccnnnnnn...
Trim 34 five prime bases, (19 Ambigs)
5' 3'
Trim 136 three prime bases, (20 Ambigs)
Quality - five (Has Chromatogram)
733 good bases, (1 Ambigs)
nngnngnntggntngnnn TCTGCCCCCCCCGGGGAC...TCTACCTGCTCCCCAACC gatnacg
Trim 18 five prime bases, (11 Ambigs)
5' 3'
Trim 7 three prime bases, (1 Ambigs)
Quality - four (Has Chromatogram)
674 good bases, (4 Ambigs)
nnnnnnggnnnnnnnnctctcctggnnn CGGGTCGAGTTCCTGGTG...CCTTATGGGCAAGGATGA ttctgatgactcacttcttctccatcc...
Trim 28 five prime bases, (17 Ambigs)
5' 3'
Trim 92 three prime bases, (3 Ambigs)
Quality - nine (Has Chromatogram)
634 good bases, (0 Ambigs)
...nnnnccttnnnnngnaggagtaggcacc ACATGAACCAGCGGGGTC...GGAGGAGCATGATGTTAT cagctccaagtcgtattctggtcgggg...
Trim 42 five prime bases, (20 Ambigs)
5' 3'
Trim 120 three prime bases, (3 Ambigs)

Sorting items

In order to assist you with the review of your data prior to trimming you can sort the fragments. Click on the **Sort Items** button at the top of the **Ends Trimming** window to order the fragments being trimmed. Figure 6-5 shows the sort criteria.

Figure 6-5 Sort Fragments dialog box



Executing the trim

Once you have reviewed your data and are happy with the criteria you have set, you proceed with the trim by clicking on the **Trim Checked Items** button. Sequencher warns you that this command cannot be reversed with the **Undo** command. To proceed, click the **Trim** button. Any red lines or bases will disappear from the **Ends Trimming** window. Now you can dismiss the window with the command from the **File** menu called **Close Window.**

Remember to save your work at this point by going to the **File** menu and using the **Save Project** command, since none of your trim work will have been recorded onto disk. Alternatively use the **Auto-Save** function to save your work at user-defined intervals (see Chapter 21 “Customizing Sequencher and User Preferences”).

Trimming vector contamination

Selecting sequences for Vector Trimming

Select one or more fragment(s) in the Project Window and go to the **Sequence** menu and choose the **Trim Vector…** command. Note that even if you select contigs, the **Trim Vector** function only affects unassembled fragments.

Specifying a vector using VecBase

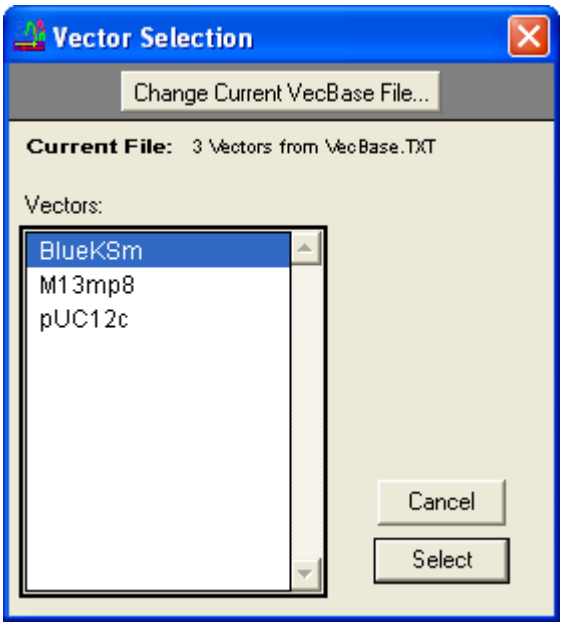
After you have selected the **Trim Vector…** command and if you have not previously specified vector insertion sites, a window will appear. Click the **Choose Insertion Site Now** button if you are ready to work with the vector trim function. Otherwise, click on **OK** to dismiss.

You can also go to the **Window** menu under **Specify Vector Insertion Sites…** to see this dialog.

Once you have pressed the **Choose Insertion Site Now** button, the **Vector Insertion Sites** dialog box appears. In the bottom part of the window you will see that you can screen for up to five vectors at once.

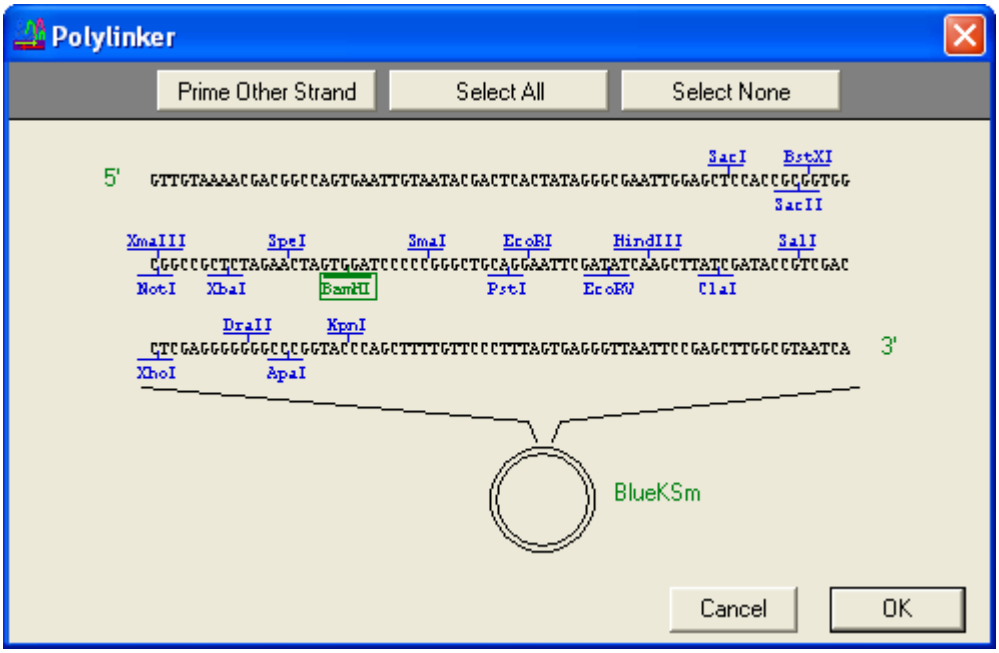
Load vector information from the VecBase database by clicking the **Use VecBase File** button in the **Vector Insertion Sites** window. Sequencher displays a list of the vectors. Select the vector you want and click the **Select** button.

Figure 6-6 Selecting a vector



The polylinker region of the vector you select is displayed in the window shown in Figure 6-7.

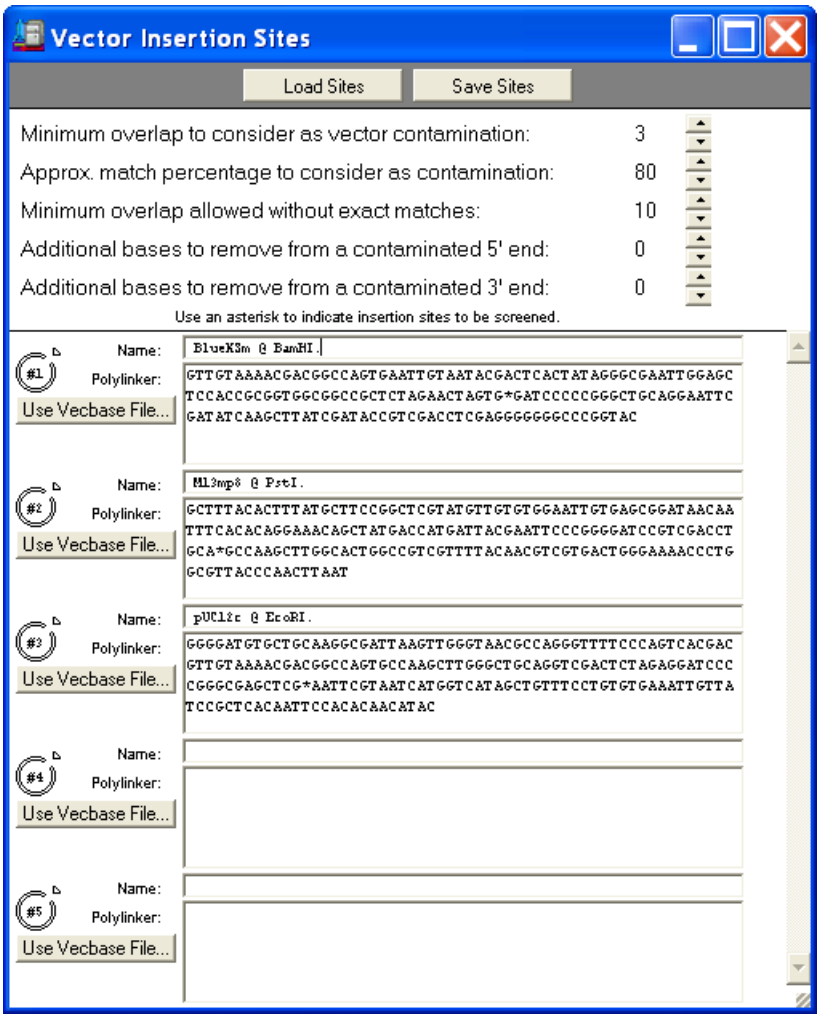
Figure 6-7 The polylinker window



Note: Not all vectors in VecBase have polylinker information. Sequencher cannot work with VecBase vectors that do not include polylinker data. Add these manually (see below).

Click the enzyme restriction site (s) where your sequence fragment was inserted into the vector. The name of a selected site(s) changes to green and is boxed. If you change your mind and decide on a different site, you must click the selected site again to deselect it. It will then return to its original color. The bases from the selected site(s) are automatically entered into the Vector dialog box, shown in Figure 6-8.

Figure 6-8 Vector Dialog box displaying vector and insertion site



If the cloning technique you are using places your insert between two sites, replace the vector sequence between the two sites with a single bullet (•), just as your insert replaces this sequence.

Specifying the insertion site manually

If the vector you select in VecBase does not have polylinker information or if you are screening for contamination using a sequence not already in VecBase, you must enter the information manually. In the **Specify Vector Insertion Sites** window paste the sequence (maximum 250 bases) into the polylinker window. Insert an "asterisk" ("*") by holding down the **Shift** key and the 8 key (not the one from the number keypad) to represent the sequence you wish to retain.

Saving and loading vector sites

You can save manually entered vector information to a file by clicking the **Save Sites** button. You can reload the file in the future by using the **Load Sites** button. This allows you to save your preferences and parameters for a later session.

Setting Vector Trimming Criteria

Sequencher has a set of default criteria to facilitate the recognition of any contaminating sequence. These criteria are listed in the top region of the **Vector Insertion Sites** window. If these parameters do not work for your specific circumstances, you can alter the vector screening parameters.

As an example, if there is ambiguity in the region of the vector junction, you can increase the likelihood that Sequencher will recognize the vector sequence by decreasing the approximate match percentage. If it is essential to remove even a single base of vector contamination, you can decrease the minimum overlap to be considered as vector contamination.

You can change any of the five independent parameters in order to improve your trim. Elevator buttons are used to increase or decrease the parameter value.

The following table describes the recognition criteria for contaminating sequence.

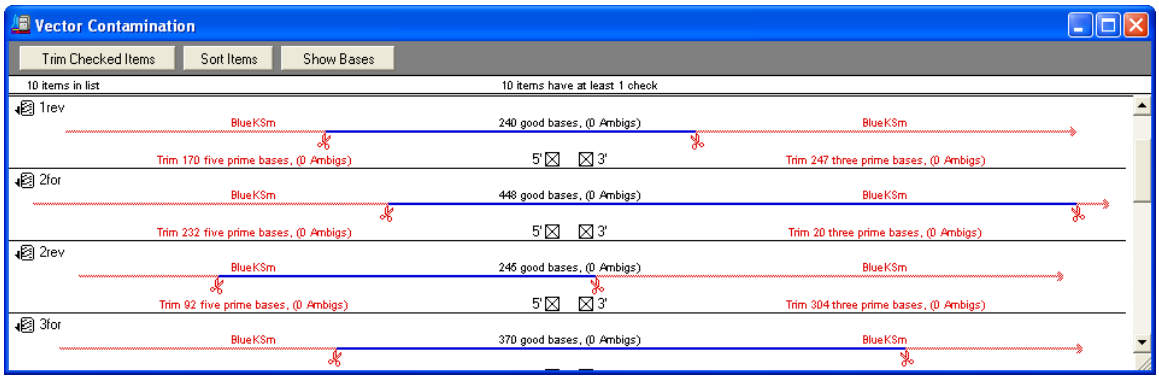
Parameter	Description
Minimum overlap to consider as vector contamination	The smallest amount of overlap to consider as vector contamination. The default value is three bases.
Approx. match percentage to consider as contamination	The percentage match Sequencher needs to find before considering the matching sequence to be vector contamination. Ambiguities do not count as full matches or mismatches.
Minimum overlap allowed without exact matches	The smallest number of matching bases within which an inexact match is permitted. If fewer bases than this match, they must match exactly.
Additional bases to remove from a contaminated 5' end	Extra material, if any, to be trimmed off the 5' end of the remaining DNA sequence along with the vector contamination. The default is zero bases.
Additional bases to remove from a contaminated 3'	Extra material, if any, to be dropped off the 3' end of the remaining DNA sequence along with the vector contamination. The default is zero bases.

Once you have finished choosing vectors and setting recognition criteria close the **Vector Insertion Sites** window.

How Sequencher screens sequences

When you execute the **Trim Vector** command, Sequencher screens the selected sequences for vector contaminants using the vector sequences you specified. Sequencher starts at the designated insertion site, the bullet, and then scans outwards. Any sequences that are found to contain vector bases are collected in the **Vector Contamination** window. (See Figure 6-9.) Please note that any contigs you have selected will be ignored, as will any sequences containing fewer than 21 bases.

Figure 6-9 The Vector Contamination window



You can print out the **Vector Contamination** window as a record for your lab notebook by going to the **File** menu and choosing **Print**. This is particularly useful if you use the **Show Bases** button to show which sequences are to be trimmed. You can also sort the window using various criteria by clicking on **Sort Items**.

Executing the trim command

To perform the vector screen and trim, go to the **Sequence** menu and choose **Trim Vector….** Once you have reviewed your data and are happy with the criteria you have set, proceed with the trim by clicking on the **Trim Checked Items** button. Sequencher warns you that this command cannot be reversed with the **Undo** command. To proceed, click the **Trim** button. Any red lines or bases will disappear from the **Ends Trimming** window. Now you can dismiss the window with the command from the **File** menu called **Close Window.**

If you wish to revert some or all of your sequences to their original state, you can go to the **Sequence** menu and click on the **Revert to Experimental Data** command.

Remember to save your work at this point by going to the **File** menu and using the **Save Project** command, since none of your trim work will have been recorded onto disk. Alternatively use the **Auto-Save** function to save your work at user defined intervals (see Chapter 21 “Customizing Sequencher and User Preferences”).

7. The Reference Sequence

In this chapter we explain Reference Sequences – sequences or fragments used as baseline or benchmarks for subsequent comparison. We will describe the special features of Reference Sequences. We will then discuss how to designate a sequence as a Reference for your project, how to assemble sequences to the Reference and how to edit contigs containing a Reference Sequence.

What is a Reference Sequence?

A Reference Sequence is any sequence that you have selected to act as a model for your subsequent comparisons. You may have obtained this sequence from one of the public DNA databases or you may have characterized it yourself. Once you have marked your sequence as a Reference, Sequencher attaches certain properties to it.

Why use a Reference Sequence?

There are many situations where you might want to designate one fragment in a project as a Reference Sequence. In forensics, human identification depends on comparing differences with a standard reference sequence. In medical genetics, a lab might want to compare the same gene from family members to the specific variant that is known to be part of a disease process. In a population study, a lab might compare the same gene from thousands or tens of thousands of individuals to a wild type sequence. In all of these circumstances, using Sequencher's **Assemble to Reference** command with a Reference Sequence allows you to pinpoint mutations and variants rapidly.

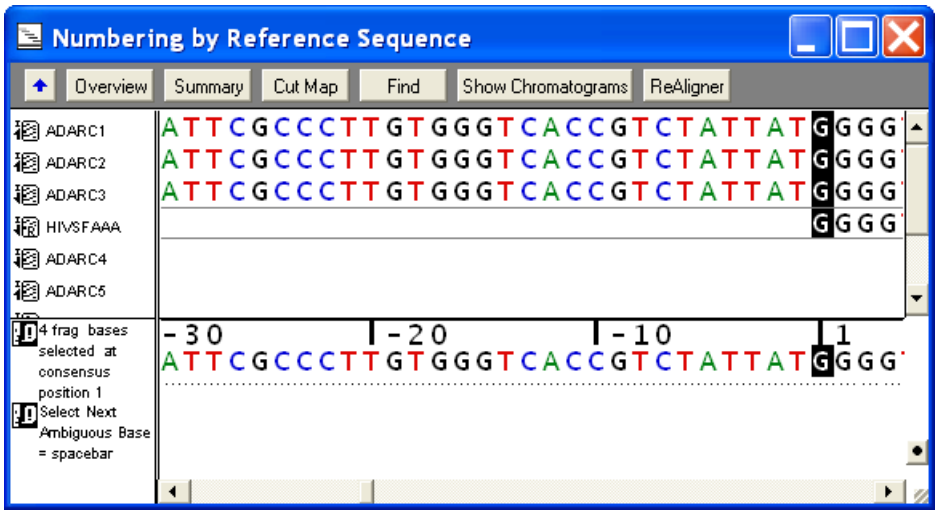
Reference Sequence properties

Before you start work with the Reference Sequence, you should be aware of a number of important properties.

Basic Reference Sequence properties

When you assemble sequences, which include a Reference Sequence, the base numbering of a contig will be the same as the base numbering of the Reference Sequence (see Figure 7-1).

Figure 7-1 Contig showing base numbering set by Reference Sequence



When a contig is built that includes a Reference Sequence, the contig will always remain in the same orientation relative to that Reference. This way you can keep the contig from *flipping* as more fragments are added. Because of this, there can only be one Reference Sequence in a contig. You can easily spot the Reference Sequence in a contig editor because it is highlighted along its entire length with a gray border.

When you assemble sequences that include a Reference Sequence you will notice that gaps in the Reference Sequence are given decimal numbers, thereby preserving the absolute numbering of base positions.

The Reference Sequence does not contribute to the calculation of the consensus line of a contig. For example, if a ‘T’ in the Reference Sequence overlaps a ‘G’ in another fragment, the consensus line will show a ‘G’.

Figure 7-2 Base numbering showing decimal numbers

Working with a Reference Sequence

How to mark a sequence as a Reference

To designate a sequence as a Reference Sequence, select the sequence icon in the Project Window and under the **Sequence** menu select the **Reference Sequence** command**.** From now on, that sequence icon will include a small letter "R" to remind you that it has been marked as a Reference.

Assemble to Reference

Assemble to Reference is a powerful command which allows you to assemble all the samples you select to a single Reference Sequence, regardless of any inconsistencies between the individual fragments. Because it is a many-to-one comparison instead of the normal many-to-many comparison, **Assemble to Reference** is much faster than the standard **Assemble Automatically** mode.

To **Assemble to Reference**, select the sequence fragments you want to assemble. Click the **Assemble to Reference** button or go to the **Contig** menu and choose the **Assemble Contigs** command, then click on **Assemble to Reference**. You will get a warning dialog box if you have not included a properly designated Reference Sequence among your selected sequences.

Note that you use the **Assembly Parameters** to change the conditions under which the assembly will take place. (See Chapter 9 "Sequence Assembly" for more information.)

To Reference by Name

To Reference by Name is a special case where the behavior of the Reference Sequence is modified to work with the **Assemble by Name** command. The **To Reference by Name** button will only be displayed if **Assemble by Name** has been enabled. (For more information on **Assemble by Name** see Chapter 9 "Sequence Assembly".)

When the **To Reference by Name** command is used, each contig created will contain a Reference Sequence and will also retain the numbering of the Reference Sequence.

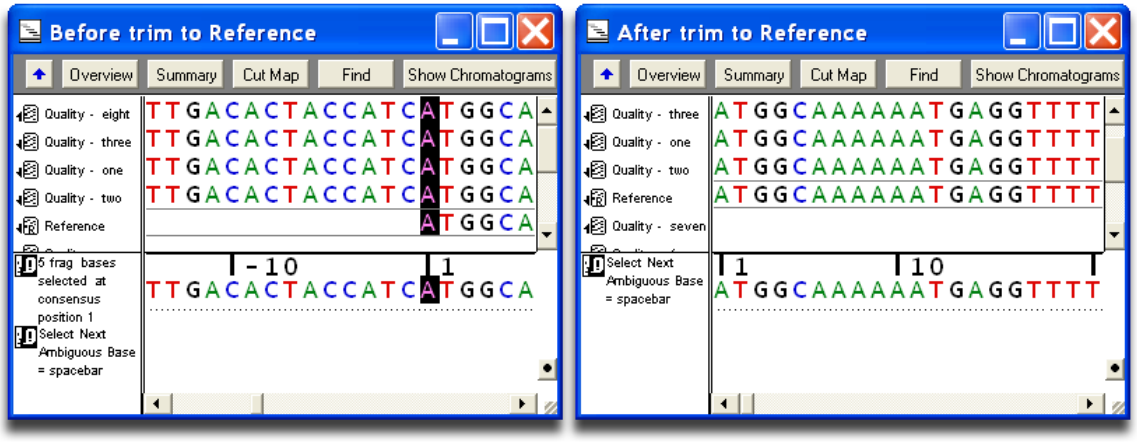
To use **To Reference by Name**, select the sequence fragments and Reference Sequence you want to assemble. Click the **To Reference by Name** button or go to the **Contig** menu and choose the **Assemble Contigs** command and then click onto **Reference by Name** from the submenu.

Trim to Reference

The **Trim to Reference Sequence** command is used to remove bases in a contig that flank the Reference Sequence on both the 5' and 3' ends (see Figure 7-3). To use this command, go to the **Contig** menu and choose **Trim to Reference Sequence**. This command is only available when you are in a contig that contains a Reference Sequence.

Note: You cannot undo this action. If there are any sequences, which do not overlap the Reference Sequence they will be removed from the contig. Sequencher will warn you which fragments have been removed in a window, which you can print for your records.

Figure 7-3 Contig with Reference Sequence flanked by overhanging sequence (before and after)



Contig Editing with a Reference Sequence

If you try to edit a Reference Sequence which has been assembled into a contig using either the standard **Assemble Automatically** or **Assemble to Reference**, a warning dialog will be displayed asking if you are sure this is what you want to do. You can proceed with the edit by clicking the **Yes** button. You can cancel the warning and continue to edit by clicking the **No** button.

The Reference Sequence is protected from any changes you make while you are editing in the consensus line. Therefore, if you delete gaps from the consensus this will not affect the Reference Sequence unless it also has gaps at the position of the deleted bases.

For example, you can delete a range of bases from all fragments by selecting a range within the **Consensus** sequence. When the deleted bases are in the middle of a fragment they will be replaced with gaps to maintain the alignment with the Reference Sequence. Since the action cannot be reverted using the **Undo** command, you must use the **Revert to Experimental Data** command from the **Sequence** menu to recover.

Fill Coverage Gaps with Reference

Under normal circumstances the Reference Sequence does not contribute to the calculation of the consensus. However, there is a special instance where the only data available in a particular region may be the data in the Reference Sequence. This might happen if you have aligned cDNAs to a genomic reference. The command **Fill Coverage Gaps with Reference** from the **Contig** menu will allow you to fill the gap with bases from the Reference Sequence. The bases from the Reference Sequence are dropped into the consensus line, but only in those places where there is no other data (Figure 7-4).

Figure 7-4 Filling gaps with the Reference Sequence

Reference Sequence Translation

This command is only available when the Reference Sequence is assembled into a contig. To see a translation of the **Reference Sequence** in the notation line below the consensus sequence, go to the **View** menu and choose **Reference Sequence Translation**. The translation has a pale gray border above and below the amino acids to distinguish it from the consensus sequence translation.

Note: As you cycle through the Consensus translation button modes, you will notice one mode which translates the consensus sequence in the same frame as the Reference Sequence. In this mode, the button icon is marked with an “r”.

8. The Sequence Editor

In this chapter we explain basic sequence editing, setting base numbers and how to reverse and complement your data. We then go onto discuss working with experimental data and how Sequencher enables you to find ambiguities, work with codon and restriction maps and annotate features. You will also learn how to customize your sequence view and use voice verification to help you check base calls.

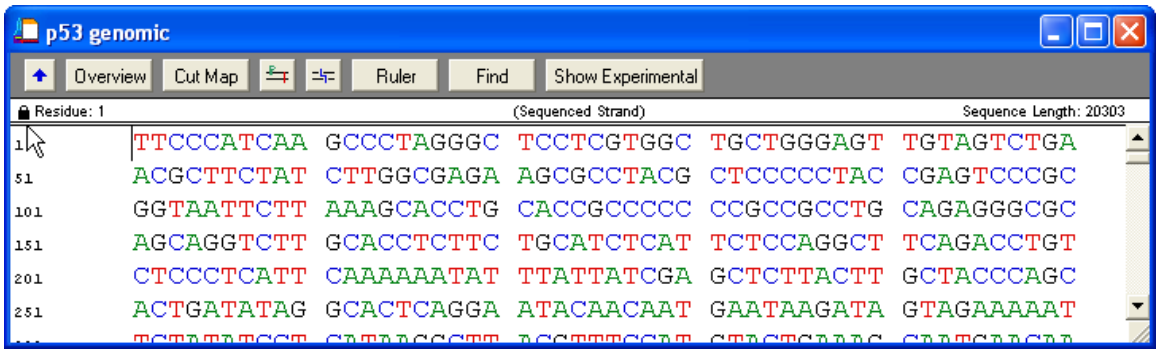
Opening an existing sequence for editing

To open an existing sequence, select a sequence icon and then go to the **File** menu and choose **Open Window**. You can also just double-click the icon to open its editor window.

Locked Editors

If an editor displays experimental data or a sequence that has been incorporated into a contig, it is locked. A padlock icon (shown to the left of the word "Residue" in the information bar above the sequence in Figure 8-1) shows that the file is locked.

Figure 8-1 A locked editor

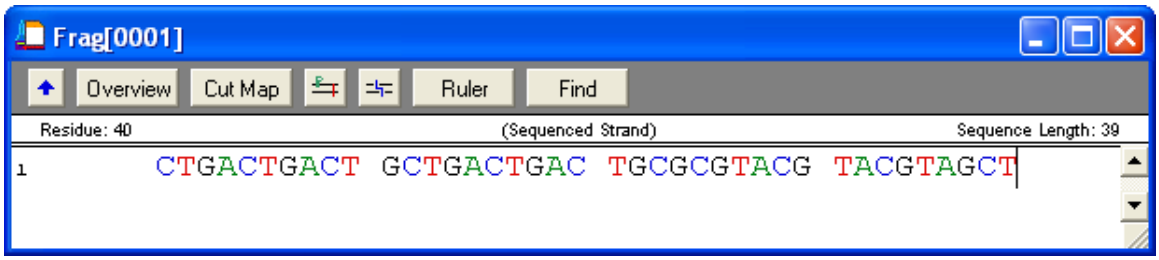


A locked editor is read-only. You cannot directly edit data, you can only view it. However, if the sequence fragment has been incorporated into a contig, you can use the contig editor to make changes. (See Chapter 12 “Editing Contigs”.)

Basic sequence editing

The Sequence Editor, shown in Figure 8-2, is a text processor that is specially optimized to work with DNA sequences.

Figure 8-2 The Sequence Editor



The default system for coding in Sequencher is the IUPAC-IUB coding system. You can define your own coding system for bases and ambiguities. (See Chapter 21, “Customizing Sequencher and User Preferences” for details. The IUPAC-IUB codes also are included in the appendix to this manual.)

The Sequence Editor works like any other text editor. The insertion point can be moved using either the mouse or the arrow keys.

Note: If you hold down the **Alt** key and use an arrow key, the insertion point moves three bases. If you hold down both the **Ctrl** and **Alt** keys, the insertion point moves nine bases. If you hold down the **Ctrl** key and the up arrow, the insertion point moves to the beginning of the sequence; if you hold down **Ctrl** key and the down arrow, the insertion point moves to the end of the sequence.

The editor also performs typical operations including **Cut, Copy, Paste** and **Undo**. Use the mouse to highlight a selection; hold down the **Shift** key and click elsewhere in the sequence with the mouse to extend the selection.

There are also two commands for extending a current selection to the beginning or end of a sequence, **To Left End** and **To Right End**, both located under the **Select** menu in the **Extend Selection** submenu.

Once you have made a selection you can copy it by going to the **Edit** menu and using the **Copy Selection** command. You can also cut the selection by using the **Cut Selection** command, also located in the **Edit** menu.

The **Paste** function will remove non-sequence characters. You can define a sequence character by going to the **Window** menu and then using the **Ambiguity Editor** in the **Ambiguity/Key Codes…** option.

For instructions on trimming ambiguous data and vectors at the time you load your fragments, see the chapters “Importing Data”, “Preparing Your Data for Assembly,” and “Customizing Sequencher and User Preferences” elsewhere in this manual.

Setting the base numbering

Open the Sequence Editor and highlight a base. Go to the **Sequence** menu and choose **Set Base Number**. From the submenu choose either **As Base 1** or **As Base Number**.

If you choose **As Base Number**, you will see a dialog box and can then enter a new number for the selected base. Click **OK** to dismiss the dialog box.

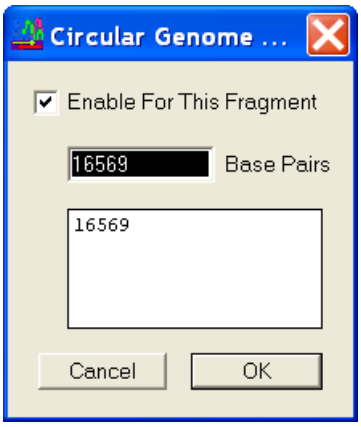
Note: If you set a base to be base 1 and this is not the first base in your sequence, every base before it will have a negative number.

Set Circular Genome Size

Sometimes you will be working with circular DNA such as plasmids or even small genomes. The **Set Circular Genome Size…** command allows you to set the number of bases in your DNA circle.

Select a sequence that has already been defined as a Reference Sequence. Then go to the **Sequence** menu and choose the **Set Circular Genome Size…** command. The dialog in Figure 8-3 below will appear. Enable the circular number by clicking the checkbox. Then enter the number of base pairs in your fragment. Dismiss the dialog window by clicking the **OK** button.

Figure 8-3 The Circular Genome Size dialog window



Note: Circular numbering can only be used in conjunction with a Reference Sequence. You must designate your sequence as a Reference Sequence before you can enable circular numbering. See Chapter 7 “The Reference Sequence” for more information.

Splitting a sequence

You can split a sequence fragment into two pieces, for instance, at a known exon boundary. Put the cursor where you want to split the sequence. Go to the **Sequence** menu and choose **Split After Selection …**. Sequencher will ask you if you want to split the sequence after the last currently selected base. Click on the **Split** button to proceed. Sequencher will split the sequence into two pieces and label the new fragments for you. If

your sequence was called “MyPUC,” it will be divided into two sequences, called “MyPUC 5' end” and “MyPUC 3' end.”

Duplicating a sequence

You can duplicate a sequence fragment. To do this, highlight the icon, then go to the **Edit** menu and choose **Duplicate Seq Fragment**. Sequencher will use the name of the original fragment as a basis for the new fragment’s name. If the fragment you want to duplicate has a chromatogram associated with it, Sequencher will ask whether you also want to duplicate the chromatogram or duplicate the sequence without chromatograms. Click on the appropriate button to proceed.

If you wish to cancel the operation, simply press the **Cancel** button. Chromatograms occupy a lot of memory but Sequencher uses a very efficient algorithm to compress trace data. However, we suggest that you only duplicate the text if you don’t need the chromatograms. (See Chapter 15 “Chromatograms” for more information.)

Reverse and Complement

The data you bring into Sequencher may not necessarily be in a particular biologically relevant orientation. To change the orientation of the sequence, go to **View** and choose **Reverse & Comp**. **Sequenced Strand.** To display the original orientation, go to the **View** menu and choose **Sequenced Strand**.

Note: You will always have two copies of your data in Sequencher, the original data and the edited version. This duplication allows you to revert to the original data, as discussed in more detail below.

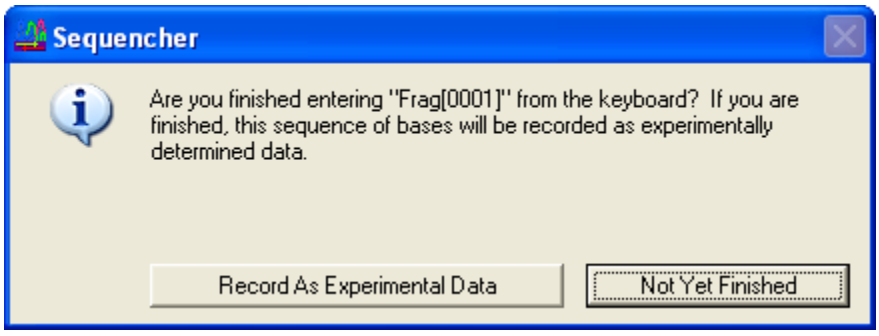
About experimental data

Creating a baseline

Sequencher notes the original base sequence when you import your data so that after editing the sequence, you can still discard all of your edits and revert to the baseline (experimental) data. You can also do this for sequences that you enter from the keyboard.

When you close a Sequence Editor that has not had its data recorded as experimental data, the program explicitly asks whether the contents of the Sequence Editor should be recorded as experimental data (see Figure 8-4). Click on **Record As Experimental Data** if you are finished editing the sequence. If you are not finished, click the **Not Yet Finished** button.

Figure 8-4 The Record as Experimental Data dialog box


Sequencher
Are you finished entering "Frag[0001]" from the keyboard? If you are finished, this sequence of bases will be recorded as experimentally determined data.
Record As Experimental Data
Not Yet Finished

Note: When a sequence becomes part of a contig the data are automatically recorded as experimental data if they have not previously been stored in that form.

Once a version of the sequence has been recorded as experimental data, all future changes are considered to be "edits" to it. As a data quality tool, all such edits can be marked in color and displayed in **Bold Magenta** or lower-case text. This feature is invoked in the **View** menu under the **Base Edits As** command. The default is set to **Bold Magenta** but you can also choose **Not Highlighted** or **bOLD & cASE cHANGE.**

Resetting a baseline

If you decide that some of the original data was incorrect (perhaps a simple typing error that has been corrected in the "edited" version of the sequence), go to the "Experimental Data" version of the sequence by clicking the **Show Experimental** button and then click the **New Baseline** button.

Note: Setting a new baseline for a sequence resets all of the bases to a not-yet-edited state. None of the bases that were edited will appear in **Bold Magenta** letters.

Viewing experimental data

After you enter a new sequence, Sequencher archives a copy that is "locked" (not editable). In the course of assembling your data, you may make a great number of edits to any particular sequence. This locked version enables you always to refer back to the sequence as it was originally entered. To see the Experimental Data version of the sequence click the **Show Experimental** button in the Sequence Editor button bar.

Reverting to experimental data

If you edit a sequence incorrectly, you can delete all the changes to that sequence and go back to the archived baseline. To do so, open the Sequence Editor . Then choose **Revert To Experimental Data** from the **Sequence** menu.

If you edit a contig incorrectly, you can delete all the changes to an individual sequence and go back to the archived baseline. To do so, open the contig editor. Select the sequence by clicking on its icon which can be found on the left hand side of the contig editor. You can now remove the sequence from the contig by using the **Remove Selected Sequences…** from the **Contig** menu. Then choose **Revert To Experimental Data** from the **Sequence** menu. You can then reselect the fragments and rebuild the contig.

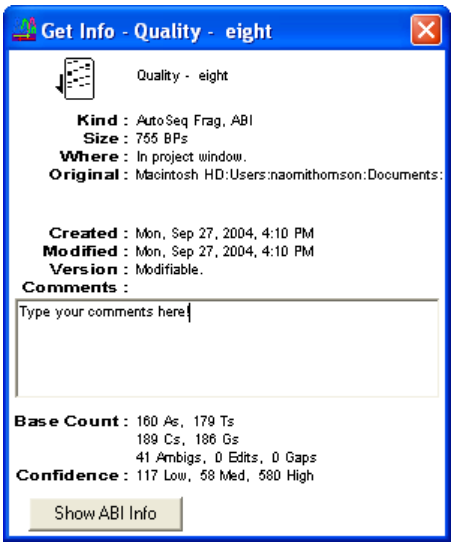
Viewing experimental data for Chromatograms

If your data came from an automated sequencer, click the button labeled **Show Chromatogram** to view the experimental data of a sequence. This will display the original trace data. Sequencher allows you to scroll either vertically (the default) or horizontally. To change the scroll orientation, click the appropriate button in the left bottom corner of the fragment chromatogram window. (See Chapter 15 "Chromatograms" for more information.)

Viewing summary information in the Sequence Editor

You can bring up summary information on your data from the Sequence Editor. Go to the **File** menu and choose **Get Info…**. The **Get Info…** window (in Figure 8-5) provides summary information on the sequence you are currently editing.

Figure 8-5 Sequence information window

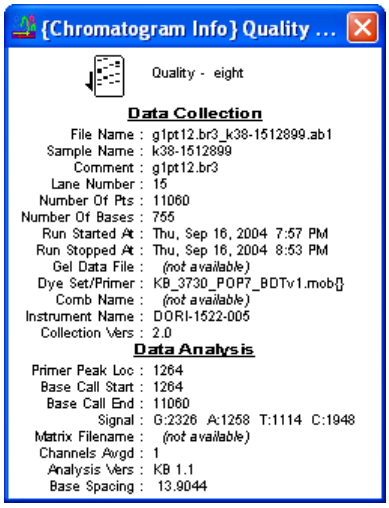

Get Info - Quality - eight
Quality - eight
Kind : AutoSeq Frag, ABI
Size : 755 BPs
Where : In project window.
Original : Macintosh HD:Users:naomithomson:Documents:
Created : Mon, Sep 27, 2004, 4:10 PM
Modified : Mon, Sep 27, 2004, 4:10 PM
Version : Modifiable.
Comments :
Type your comments here
Base Count : 160 As, 179 Ts
189 Cs, 186 Gs
41 Ambigs, 0 Edits, 0 Gaps
Confidence : 117 Low, 58 Med, 580 High
Show ABI Info

If you want to annotate your data with more information than the comment box allows you to see, go to the **Edit** menu and click on **Edit Comments…** Sequencher will open a small text editing window that stores about 250 characters of text with each sequence. After you have added text with **Edit Comments**, a small "I" (for "Info") will appear at the upper left corner of the icon.

Base counts are displayed at the bottom of the window. If your sequence also includes confidence scores you will also see a base count for each of the three confidence ranges.

If you use an ABI sequencing system click the **Show ABI Info** button. Sequencher displays a window with information from the ABI data sheet, shown in Figure 8-6.

Figure 8-6 ABI data sheet info


{Chromatogram Info} Quality ...
Quality - eight
Data Collection
File Name : g1pt12.br3_k38-1512899.ab1
Sample Name : k38-1512899
Comment : g1pt12.br3
Lane Number : 15
Number Of Pts : 11060
Number Of Bases : 755
Run Started At : Thu, Sep 16, 2004 7:57 PM
Run Stopped At : Thu, Sep 16, 2004 8:53 PM
Gel Data File : (not available)
Dye Set/Primer : KB_3730_POP7_BDTv1.mob{}
Comb Name : (not available)
Instrument Name : DORI-1522-005
Collection Vers : 2.0
Data Analysis
Primer Peak Loc : 1264
Base Call Start : 1264
Base Call End : 11060
Signal : G:2326 A:1258 T:1114 C:1948
Matrix Filename : (not available)
Channels Avgd : 1
Analysis Vers : KB 1.1
Base Spacing : 13.9044

Selecting Bases

Sequencher offers a number of ways to select bases. You can choose commands from the **Select** menu or you can use a keyboard combination. To select bases in the opposite orientation, press and hold the **Shift** key when using any **Select Next…** command (from the menu or with a keyboard shortcut).

Finding Ambiguous Bases

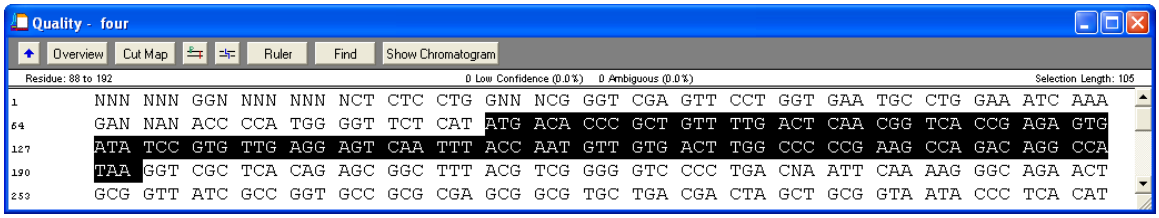
One of the most frequently executed commands is **Next Ambiguous Base** under the **Select** menu. You can use this command in the Sequence Editor or the contig editor. By using this command in the consensus sequence of a contig, Sequencher will display the next position in your contig that contains an ambiguity or disagreement. The shortcut for this command is **Ctrl+N** (for more on keyboard shortcuts see Appendix 23 “Keyboard shortcuts”). If you use the **Shift** key with this key combination, you will go back to the previous disagreement.

Once you have used any of the **Select Next** commands, Sequencher will activate the space bar to execute that operation. If you use either the menu command or the shortcut key combination two consecutive times, Sequencher will remind you that this easier alternative is available.

Finding Open Reading Frames (ORF)

Choose **Next Met to Stop (>0b)** to highlight the next pair of start and stop codons in one of the three forward reading frames. The number of bases shown in this command depends on whether you specified a preference for a minimum length in the **User Preferences**. (For instructions on how to set a specified minimum length for open reading frames see Chapter 21 “Customizing Sequencher and User Preferences”.) If you set preferences to highlight ORFs only when they are of some minimum length, the menu command will display the number of bases you specified.

Figure 8-7 ORF selected by Next Met to Stop command

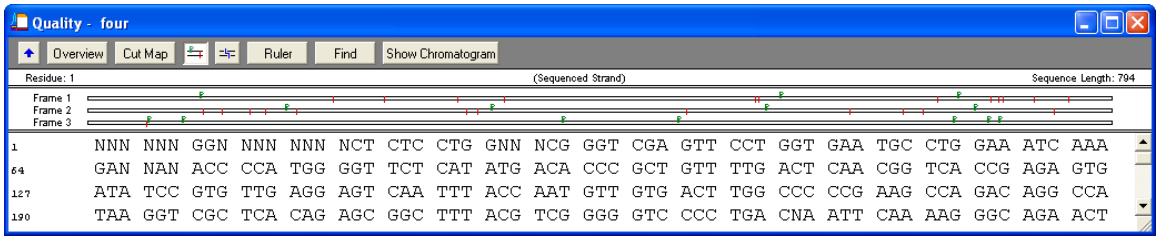

Quality - four
Overview
Cut Map
Ruler
Find
Show Chromatogram
Residue: 88 to 192
0 Low Confidence (0.0%)
0 Ambiguous (0.0%)
Selection Length: 105
1 NNN NNN GGN NNN NNN NCT CTC CTG GNN NCG GGT CGA GTT CCT GGT GAA TGC CTG GAA ATC AAA
64 GAN NAN ACC CCA TGG GGT TCT CAT ATG ACA CCC GCT GTT TTG ACT CAA CGG TCA CCG AGA GTG
127 ATA TCC GTG TTG AGG AGT CAA TTT ACC AAT GTT GTG ACT TGG CCC CCG AAG CCA GAC AGG CCA
190 TAA GGT CGC TCA CAG AGC GGC TTT ACG TCG GGG GTC CCC TGA CNA ATT CAA AAG GGC AGA ACT
253 GCG GTT ATC GCC GGT GCC GCG CGA GCG GCG TGC TGA CGA CTA GCT GCG GTA ATA CCC TCA CAT

Sequence features

Looking at codon maps

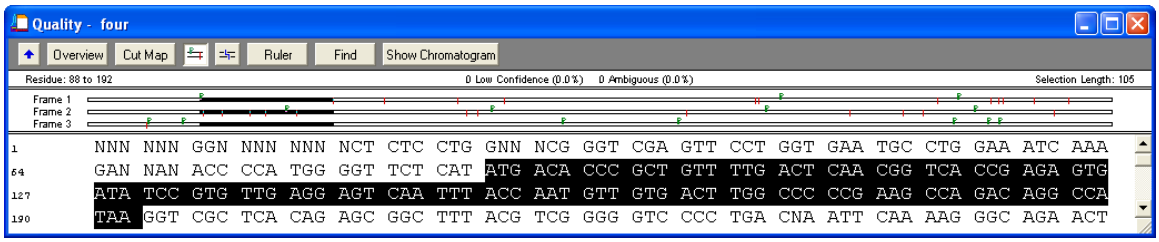
Click the **Codon Map** button located in the button bar at the top of the Sequence Editor or go to the **View** menu and choose **Display Codon Map**. The three "outlined bars" at the top of the editor window represent the three forward frames of translation. Each green "flag" sticking partially into the top of a frame marks the position of a start codon, while each red line cutting through the frame and extending below marks the position of a stop codon. Clicking between any two markers highlights that section of the sequence.

Figure 8-8 Sequence Editor with codon map button selected


Quality - four
Overview
Cut Map
Ruler
Find
Show Chromatogram
Residue: 1
(Sequenced Strand)
Sequence Length: 794
Frame 1
Frame 2
Frame 3
1 NNN NNN GGN NNN NNN NCT CTC CTG GNN NCG GGT CGA GTT CCT GGT GAA TGC CTG GAA ATC AAA
64 GAN NAN ACC CCA TGG GGT TCT CAT ATG ACA CCC GCT GTT TTG ACT CAA CGG TCA CCG AGA GTG
127 ATA TCC GTG TTG AGG AGT CAA TTT ACC AAT GTT GTG ACT TGG CCC CCG AAG CCA GAC AGG CCA
190 TAA GGT CGC TCA CAG AGC GGC TTT ACG TCG GGG GTC CCC TGA CNA ATT CAA AAG GGC AGA ACT

In the example shown in Figure 8-9, a click between a start and a stop codon of the first reading frame has selected the corresponding bases in the editor. Note that **User Preferences** have been set to highlight ORFS of more than 100 bases. There is one such ORF in Frame 2, marked with a hash pattern. To dismiss the map click on the **Codon Map** button again.

Figure 8-9 Sequence Editor with codon map

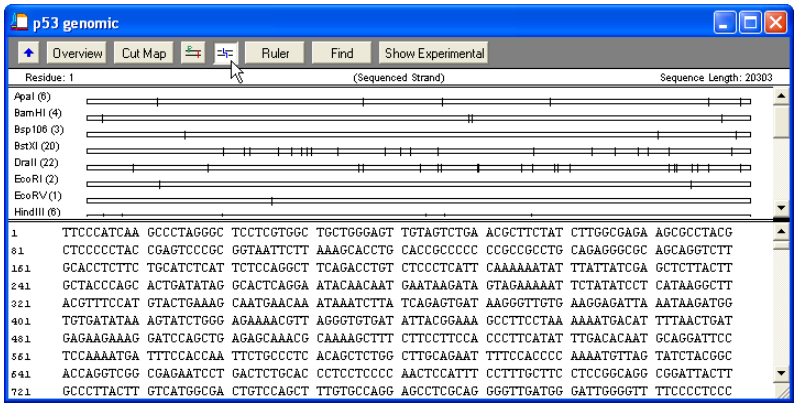

Quality - four
Overview
Cut Map
Ruler
Find
Show Chromatogram
Residue: 88 to 192
0 Low Confidence (0.0%)
0 Ambiguous (0.0%)
Selection Length: 105
Frame 1
Frame 2
Frame 3
1 NNN NNN GGN NNN NNN NCT CTC CTG GNN NCG GGT CGA GTT CCT GGT GAA TGC CTG GAA ATC AAA
64 GAN NAN ACC CCA TGG GGT TCT CAT ATG ACA CCC GCT GTT TTG ACT CAA CGG TCA CCG AGA GTG
127 ATA TCC GTG TTG AGG AGT CAA TTT ACC AAT GTT GTG ACT TGG CCC CCG AAG CCA GAC AGG CCA
190 TAA GGT CGC TCA CAG AGC GGC TTT ACG TCG GGG GTC CCC TGA CNA ATT CAA AAG GGC AGA ACT

Looking at Restriction Maps

You can look at a restriction map for the sequence by clicking the **Cut Map** button at the top of the Sequence Editor. A restriction map replaces the sequence in the window. (See Chapter 16, "Restriction Maps," for a general discussion of restriction maps.)

Sequencher can add a pane showing the restriction map for a sequence into the same window as the sequence itself. At the top of the Sequence Editor is a button showing an icon for a staggered enzyme cut (shown in Figure 8-10).

Figure 8-10 Restriction Map button selected with map and sequence shown



Click the button; a window with a multi-line restriction map appears. You can adjust the size of the panes by dragging the splitter bar that separates the scroll bars for each pane (see Figure 8-10). You can also invoke this command by going to the **View** menu and choosing **Display Cut Map Inset.**

Like the full-page cut map, the window restriction map is linked to the Sequence Editor. If you click between two cut sites, Sequencher highlights the resulting restriction fragment, as shown in Figure 8-11. If you click on the name of an enzyme, Sequencher highlights the recognition sites.

Figure 8-11 Marking restriction and recognition sites

Annotating Sequences

You can assign a standardized GenBank **Feature Key** in addition to personal feature annotations to describe subsequences. Select the range of bases you wish to annotate, then choose **Mark Selection As Feature** from the **Sequence** menu.

If you want to assign a personal annotation, select **Sequencher** from the drop down menu called **Feature Key:.** To give the feature a name, type your text into the **Feature Name:** box**.** You can use the default display style or you can assign a **Feature Color** and **Feature Style**, such as inverted case or underlined, from the dialog box. You can also use the **Display:** radio buttons to determine whether single stranded DNA is displayed or RNA, protein translation or the complement will be shown (For more information see Chapter 17 “Motifs and Features”.)

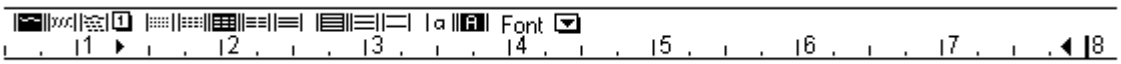
If you want to assign a GenBank feature key, make a selection from the drop down menu called **Feature Key.** To give the feature a name, type your text into the **Feature Name:** box**.** You can use the default display style or you can assign a **Feature Color** and **Feature Style**, such as inverted case or underlined, from the dialog box. You can also use the **Display:** radio buttons to determine whether single stranded DNA is displayed or RNA, protein translation or the complement will be shown (For more information see Chapter 17 “Motifs and Features”.)

Customizing your sequence view

Formatting Ruler

Formatting can be controlled with a ruler (shown in Figure 8-12) that lets you adjust several display parameters. Click the **Ruler** button in the Sequence Editor or contig Summary Window. Alternately you can go the **View** menu and choose **Display Format Ruler.**

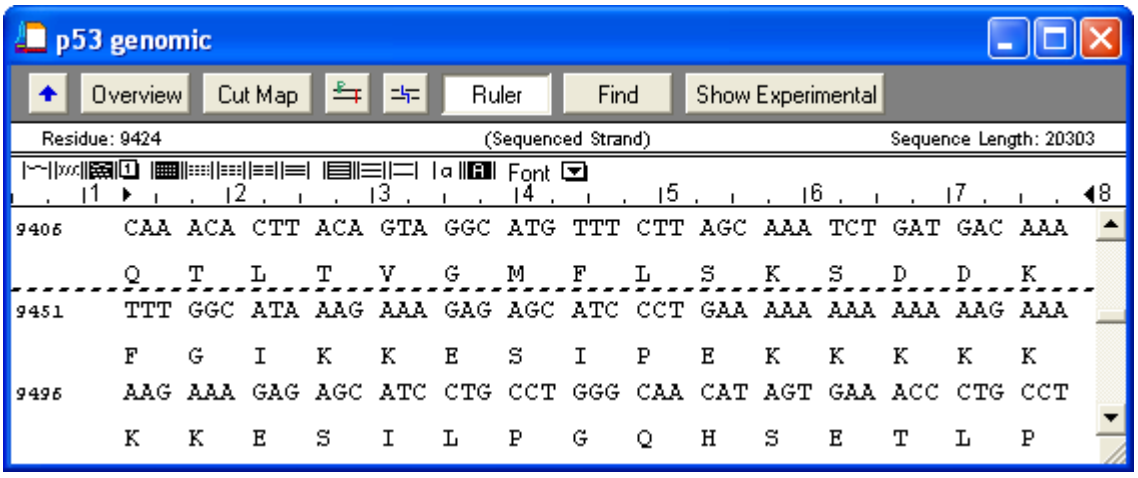
Figure 8-12 The formatting ruler



Margins

You can drag the margin triangles on this ruler to set the width of your sequence. Figure 8-13 shows margins that are limited to 42 bases per line.

Figure 8-13 Sequence Editor with ruler and translations


p53 genomic
Overview
Cut Map
Ruler
Find
Show Experimental
Residue: 9424
(Sequenced Strand)
Sequence Length: 20303
Font
9406 CAA ACA CTT ACA GTA GGC ATG TTT CTT AGC AAA TCT GAT GAC AAA
Q T L T V G M F L S K S D D K
9451 TTT GGC ATA AAG AAA GAG AGC ATC CCT GAA AAA AAA AAA AAG AAA
F G I K K E S I P E K K K K K
9496 AAG AAA GAG AGC ATC CTG CCT GGG CAA CAT AGT GAA ACC CTG CCT
K K E S I L P G Q H S E T L P

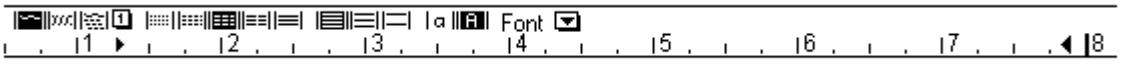
The sequence margins you set with the formatting ruler are not the same as your absolute print margins, which allow room at the left to show the first base number of a line. To change the absolute limits, go to the **File** menu and choose **Set Header & Footer…** and enter new values into the **Margin** boxes or drag the triangular markers just under the Translation and Base Grouping icons, as shown in Figure 8-13 above.

Translation

On the left, just under the word "Residue" and above the ruler, is a set of three small icons that let you specify sequence representation: single-stranded, double-stranded, or single-stranded with a protein translation.

If you select the third 'translation' icon, the numbered button at the right controls which frame or whether all three frames will be displayed. You can also use the translation commands in the **Translation** submenu of the **View** menu: **Single Stranded**, **Double Stranded**, **Protein 1st Frame, Protein 2nd Frame**, **Protein 3rd Frame**, or **Protein All 3 Frames** (see Figure 8-14).

Figure 8-14 Translation icons on the Formatting ruler



See Chapter 21, "Customizing Sequencher and User Preferences" for information on editing the genetic code or using one-letter abbreviations for amino acids.

Base grouping

To the right of the translation display icons are the icons that let you choose blocking, the number of bases to be displayed in a single group. The available settings are **3**, **5**, **10**, **20** and **Fill-to-margin**. In Figure 8-14 above, the icon for a grouping of 10 is selected.

Line spacing

You can choose single, double or triple line spacing. The three small icons that let you control the line spacing are located to the right of the base grouping icons. Figure 8-14 above shows the icon for single spacing selected.

Case and Font

To the right of the line spacing icons is a pair of icons that let you choose upper or lower case for the display. In Figure 8-14 above, the icon for upper case is selected.

To the right of the upper and lower case icons is a pull down menu, which lets you choose a font and size. Only fixed-width fonts are displayed, ensuring that the rows of bases will line up vertically.

Voice verification

Sequencher provides two forms of audio feedback to help you maintain the quality of data input when you are entering information manually from the keyboard. You can have Sequencher announce each base as you type it in. You can have the computer "read" the sequence aloud back to you while you look at the original source of the data. Sequencher stores the sounds that it uses in a sound file stored in your Sequencher folder.

Selecting a voice file

First, make sure your computer sound system is enabled and that the volume is not set on mute.

Then, go to the **Sequence** menu and choose **Speech**. Then go to the submenu **Select Voice File….**You will not get any audio feedback until the **Select Voice** file is loaded.

Identifying a speaker

You'll probably be familiar with the voices Gene Codes has persuaded to provide audio for you! To hear whose voice is being used, go to the **Sequence** menu and choose **Speech** and then the **Identify Speaker** subcommand. You can only select this command if a sound file was selected previously.

Audible keystrokes

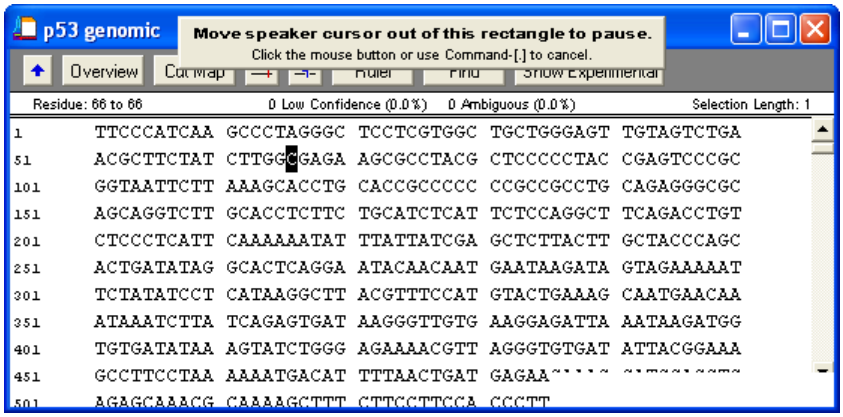
Any meaningful keystroke you type into a Sequence Editor will be spoken aloud if you go to the **Speech** command in the **Sequence** menu and select the **Say Sequence Keystrokes** submenu. That is, if you press the "A" key, a voice will say "A," but if you press the shift key, you will not hear any voice.

Read sequence selection

To check the sequence you have entered into the Sequence Editor, first select the bases you want the computer to read to you. Then go to the **Sequence** menu, click on **Speech** and then the **Read Sequence Selection** submenu.

Note: You must place the cursor somewhere in the small window that appears above the sequence as in Figure 8-15. To pause the reading, just move the cursor out of the window.

Figure 8-15 Speak selection cursor and instruction menu


p53 genomic
Move speaker cursor out of this rectangle to pause.
Click the mouse button or use Command-[.] to cancel.
Overview
Residue: 66 to 66
0 Low Confidence (0.0%)
0 Ambiguous (0.0%)
Selection Length: 1
1 TTCCCATCAA GCCCTAGGGC TCCTCGTGGC TGCTGGGAGT TGTAGTCTGA
51 ACGCTTCTAT CTTGGCGAGA AGCGCCTACG CTCCCCCTAC CGAGTCCCGC
101 GGTAATTCTT AAAGCACCTG CACCGCCCCC CCGCCGCCTG CAGAGGGCGC
151 AGCAGGTCTT GCACCTCTTC TGCATCTCAT TCTCCAGGCT TCAGACCTGT
201 CTCCCTCATT CAAAAAATAT TTATTATCGA GCTCTTACTT GCTACCCAGC
251 ACTGATATAG GCACTCAGGA ATACAACAAT GAATAAGATA GTAGAAAAAT
301 TCTATATCCT CATAAGGCTT ACGTTTCCAT GTACTGAAAG CAATGAACAA
351 ATAAATCTTA TCAGAGTGAT AAGGGTTGTG AAGGAGATTA AATAAGATGG
401 TGTGATATAA AGTATCTGGG AGAAAACGTT AGGGTGTGAT ATTACGGAAA
451 GCCTTCCTAA AAAATGACAT TTTAACTGAT GAGAA
501 AGAGCAAACG CAAAAGCTTT CTTCCTTCCA CCCTT

To speed up or slow down the rate at which Sequencher's audio "reads" your data, go to the **Window** menu and choose **User Preferences...**. Under the **General** section you will find an item called **Sound**. Click on this to display a slider bar which controls the speed of read back in bases per second. To end the read back, just click your cursor anywhere on the Sequence Editor window.

9. Sequence assembly

In this chapter you will learn how to set assembly conditions by using Sequencher's rich variety of assembly algorithms and parameters. You will learn about automatic assembly, assembling interactively, assembling groups of sequences using sequence name information, the large gap algorithm for assembling cDNA and genomic DNA and assembling to Reference Sequences. You will learn how to set assembly conditions for each of the different algorithms.

The Assembly Strategy

Sequencher provides you with powerful options for performing your sequence assembly. In many cases the default options will perform well. However, experimental variation can affect your data, so there will be times when you may need to change the assembly parameters. The following is a basic strategy for assembling your data.

- Start with strict parameters to minimize the possibility of incorrect matches. Select the sequences you want to assemble and use the **Assemble Automatically** option. Automatic assembly relies on an exhaustive search-and-compare algorithm.
- If some sequences do not assemble, decrease the stringency of the parameters and try **Assemble Automatically** again.
- If the selected parameters become relaxed enough that the **Assemble Automatically** process may result in scientifically insupportable alignments, you should switch to interactive alignment so you can exert more control over the process.

The most frequently used options; **Assemble Automatically**, **Assemble Interactively**, and **Assemble to Reference** are available as prominent buttons in the Project Window. In addition **Auto Assemble by Name** and **To Reference by Name** replace **Assemble Automatically** and **Assemble to Reference** when **Assemble by Name** is enabled (see

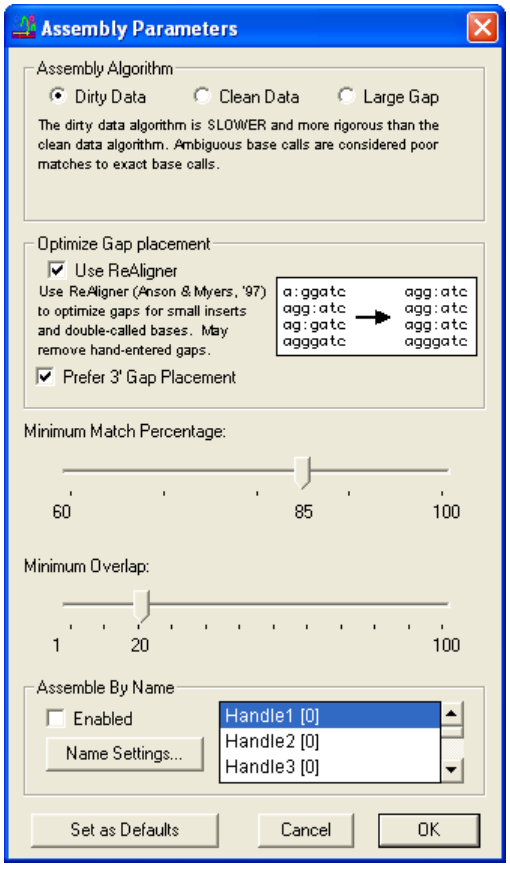
Chapter 9 “Sequence Assembly” for more information). There are numerous other options under the **Assembly Parameters** and the **Contig** menu.

Setting the Assembly Conditions

Setting assembly parameters

To change the parameter settings, select the Project Window. Then either click the **Assembly Parameters** button in the Project Window or go to the **Contig** menu and choose the **Assembly Parameters** command. Sequencher displays the dialog box shown in Figure 9-1.

Figure 9-1 Assembly parameters dialog box



Sequencher uses the values set in this window to control the way it assembles sequences. Once you have set the options you require click the **OK** button to dismiss the **Assembly Parameters** window.

Assembly Algorithms

To choose which assembly algorithm you want to use, click on the **Assembly Parameters** button on the Project Window and then click on one of the three radio buttons at the top of the window. Your choice will depend on your data.

For instance if you click the **Clean Data** radio button, contig assembly is faster but Sequencher may miss some possible matches if the ends of your sequences contain a large number of ambiguities. For best performance, make sure that you have trimmed poor quality data from your sequences.

The **Dirty Data** algorithm is somewhat slower because Sequencher is performing more rigorous comparisons between the sequences. You should note that a*utomated sequencers create dirty data* and Sequencher's algorithm is optimized to deal with that.

The **Large Gap** algorithm allows you to assemble sequences that are expected to contain insertions and deletions (gaps) greater than 10 bases long. Such gaps are typically found in evolutionary studies and sequence comparisons from different organisms. The **Large Gap** algorithm is similar to and slower than the dirty data algorithm but allows larger gaps to occur when performing the assembly. Typical examples of assemblies that require the **Large Gap** algorithm include the comparisons of a cDNA sequence to a genomic sequence and the assembly of related genes with alternative splicing.

Assembling with Dirty Data

Go to the **Assembly Parameters** dialog box and select the **Dirty Data** radio button. There are two sliders for changing minimum match values. To set the *proportion* of bases that have to match use the **Minimum Match Percentage** slider. To set the minimum number of bases that must overlap use the **Minimum Overlap** slider.

Change the settings by positioning your cursor on the slider, holding the mouse button down and dragging the slider left or right. The value set will automatically update as you move the slider.

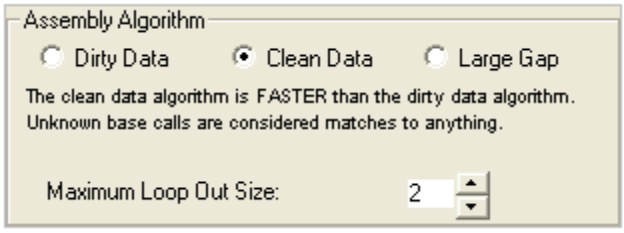
For example, if you attempt to assemble a 17-base sequence that matches perfectly to a selected 1000 base sequence with the minimum overlap set at 20 bases (Sequencher's default minimum overlap), the two fragments will not assemble. You must first reduce minimum overlap to 17 bases or lower and then assemble the fragments.

Assembling with Clean Data

When you select the **Clean Data** algorithm, an additional option called **Maximum Loop Out Size** is available. This has an elevator button for adjusting size value (Figure 9-2 below). This parameter is the largest number of *consecutively* mismatched bases that are acceptable in a potential overlap. Change the setting by positioning your cursor on the appropriate arrowhead, and holding the mouse button down. The maximum value, which can be set, is 6.

To set the *proportion* of bases that have to match use the **Minimum Match Percentage** slider. To set the minimum number of bases that must overlap use the **Minimum Overlap** slider.

Figure 9-2 The Clean Data algorithm with Maximum Loop Out Size shown


Assembly Algorithm
Dirty Data
Clean Data
Large Gap
The clean data algorithm is FASTER than the dirty data algorithm.
Unknown base calls are considered matches to anything.
Maximum Loop Out Size:
2

Assembling with Large Gap

The **Large Gap** algorithm is specifically for data that has or might have large insertions or gaps. This might include data from evolutionary studies, different organisms or comparing cDNA to genomic DNA. When you select the **Large Gap** algorithm, there are two sliders for changing minimum match values. To set the *proportion* of bases that have to match use the **Minimum Match Percentage** slider. To set the minimum number of bases that must overlap use the **Minimum Overlap** slider.

Not:e **Assemble to Reference** cannot be used with the **Large Gap** algorithm.

Assemble By Name

You can enable **Assemble By Name** by clicking in the checkbox called **Enabled**. Click the **Names Settings...** button. Choose your delimiter from the **Name Delimiters** drop down menu. Then type names for the **Assembly Handles** into the fields in the **Assembly Handles** pane.

To activate an Assembly Handle, click the adjacent radio button. When you are done, click the **OK** button.

You will notice that the drop down menu in the Assemble By Name pane under the Assembly Parameters window now contains the names of your handles.

Next to each handle is a number enclosed in square brackets. This indicates the probable number of contigs you can generate using that handle.

If you want to use a more complex delimiter, you must select **Advanced Expression…** from the Name Delimiters drop down menu. Clicking the **Define…** button will take you to the Advanced Expression for Name Parsing window. In this window, you write and preview the regular expression you want to use. You can find more information on regular expressions by going to Appendix 24 “Advanced Expressions,” or for more detail, consult the tutorial in the Sequencher Tutorials folder.

You can find more information on **Assemble by Name** in Chapter 10 “Assemble by Name”. For information on regular expressions see Appendix 24 “Advanced Expressions,” or for more detail, consult the tutorial in the Sequencher Tutorials folder.

Refining the Assembly conditions

Minimum Match Percentage

The **Minimum Match Percentage** can be used with all three assembly algorithms. It is used to set the proportion of bases which must match in candidate sequences before Sequencher accepts the sequences as actually overlapping. The default value of 85% can be changed by moving a slider.

If this value does not give you any overlaps you may need to decrease the **Minimum Match Percentage** by moving the slider to the left to lower the percentage.

If you have an incorrect or unexpected overlap you may need to increase the **Minimum Match Percentage** by moving the slider to the right (higher percentage).

Minimum Overlap

The **Minimum Overlap** can be used with all three assembly algorithms. It is used to set the minimum number of bases that must overlap before Sequencher accepts the sequences as actually overlapping. The default value of 20 can be changed by moving a slider.

If this value does not give you any overlaps you can decrease the **Minimum Overlap** by moving the slider to the left to indicate a smaller number of bases that must overlap**.**

If you have an incorrect or unexpected overlap you may need to increase the **Minimum Overlap** by moving the slider to the right (larger number of bases that must overlap).

Maximum Loop Out Size

The **Maximum Loop Out Size** is the largest number of consecutively mismatched bases (in this instance a gap also counts as a mismatch) that are acceptable in a potential overlap. The number of mismatching bases is set by moving an elevator button up or down. The maximum value that can be set manually is 6. Remember that this option can be used only with the **Clean Data** algorithm.

Optimization of gap placement

ReAligner is an optional step which can be used with the **Clean Data** and **Dirty Data** algorithms. It evaluates the disposition of gaps within a contig and optimizes their placement**. ReAligner** facilitates editing in the consensus sequence and clearly displays the effect of insertions and deletions.

To use **ReAligner** select the checkbox in the **Assembly Parameters** dialog box. In some areas of DNA analysis, the standard is to gather the gaps to the right. If you require this, select **Prefer 3' Gap Placement**.

Note: This option cannot be used with the **Large Gap** algorithm.

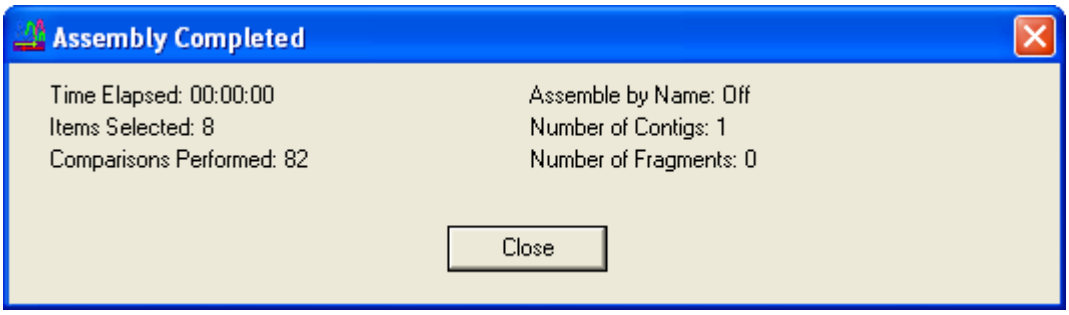
Performing the assembly

Automatic assembly

Bring the Project Window to the front and select the sequence fragments you want to assemble. Click the **Assemble Automatically** button or go to the **Contig** menu, choose the **Assemble Contigs** command and **Automatically** from the submenu. Sequencher compares all the selected sequences including reverse complements and assembles the best matches that fall within the chosen assembly parameters.

When assembly is done, the advisory window appears. (See Figure 9-3.) Click **Close** to go back to the Project Window.

Figure 9-3 Assembly done alert



Adding selected items to others – incremental building of contigs

The assembly option **Add Selected Items To Others**, which is a special case of automatic assembly, is very useful when you have a growing data set. Sequencher compares all the selected (new) items to the unselected (old) items. It will also compare the new items to each other. It has been designed to increase the efficiency of handling large numbers of sequences in efforts such as DNA library clustering and genome assembly. Therefore, it will not compare any unselected items to each other again.

Go to the **Contig** menu and click on **Add Selected Items To Others** and then on **Assemble Contigs.**

Assemble to Reference

The **Assemble to Reference** command allows you to assemble samples to a single Reference Sequence regardless of inconsistencies between the individual fragments. Because it is a many-to-one comparison instead of the normal many-to-many comparison, **Assemble to Reference** is much faster than the standard **Assemble Automatically** algorithm.

To use this command you need to have a Reference Sequence already designated. Click on your sequence and then go to the **Sequence** menu and select **Reference Sequence**. Now select the sequence fragments you want to assemble, including the Reference Sequence. Click the **Assemble to Reference** button or choose the **Assemble Contigs** command from the **Contig** menu and **Assemble to Reference** from the submenu.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Interactive assembly

The **Assemble Interactively** algorithm can be used whenever you want complete control over the assembly of a batch of sequences. The **Assemble Interactively** window provides you with detailed information regarding overlap, mismatch and gapping for any given sequence and a set of candidate sequences. The candidate assemblies are driven by your user-defined assembly parameter settings.

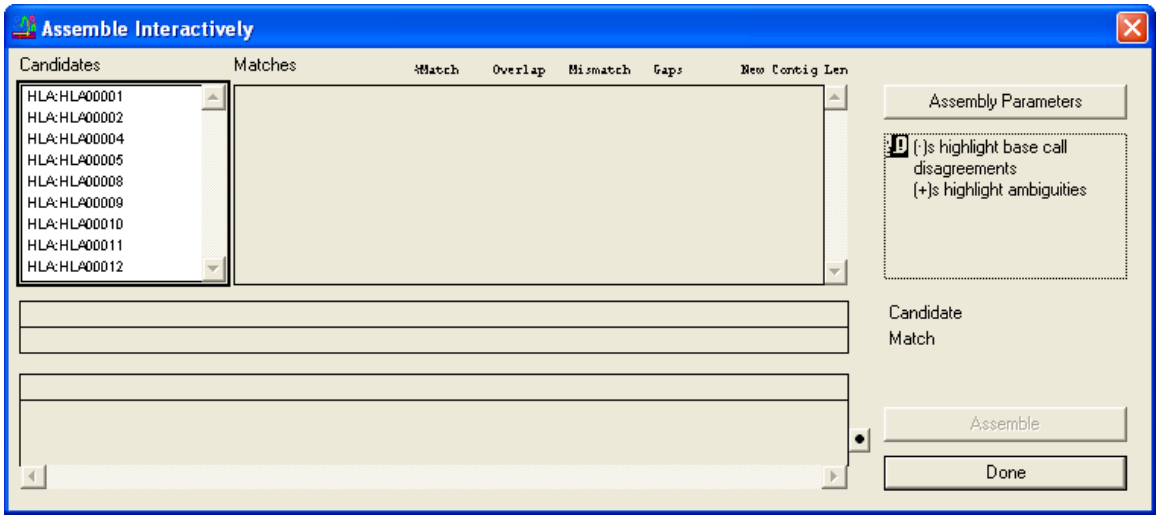
Performing an interactive assembly

Bring the Project Window to the front and select the items you want to assemble. Click on the **Assemble Interactively** button at the top of the Project Window or go to the **Contig** menu and choose **Assemble Contigs** and then **Interactively…** from the submenu.

Sequencher displays the **Assemble Interactively** dialog box shown in Figure 9-4. The sequences and contigs you selected are displayed in the **Candidates** list. The panel on the right, showing a face in the upper left corner, is called the "agent." It displays information about the comparisons you have requested. In the center is the **Matches** panel which displays the possible matches. For each match in the panel Sequencher lists the **% Match**, the approximate number of **Overlap** bases, the number of **Mismatch** bases, the number of **Gaps** and the **New Contig Length.**

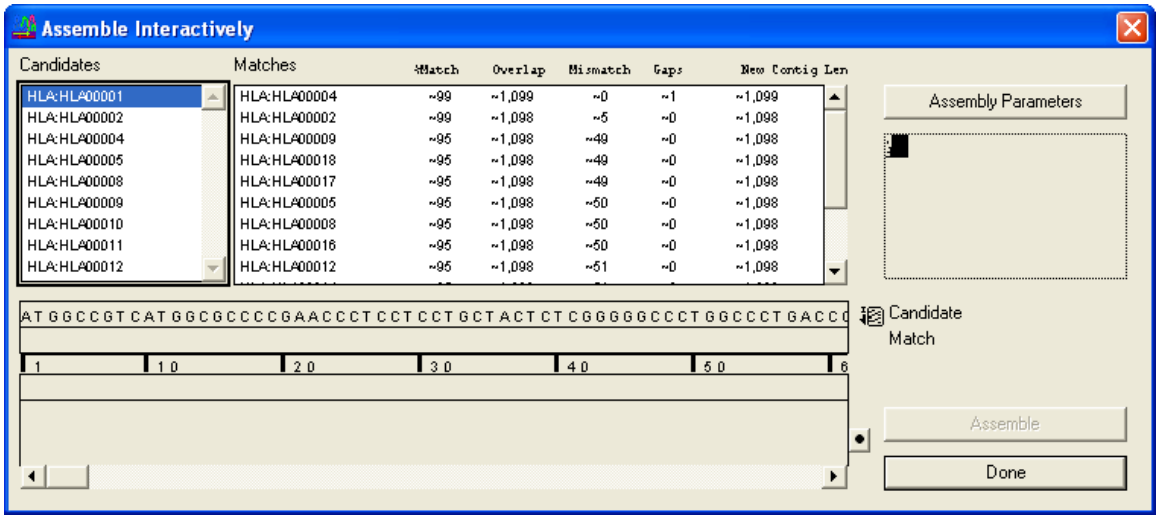
The top line shows the Candidate sequence and the line below this shows the Match sequence. Below these two lines is the consensus sequence. There is a small button to the right of the consensus labeled with a black circle. If you click repeatedly on this button you can display the first, second, third reading frame or all three reading frames at once. The label on the button changes to reflect which of these choices is active.

Figure 9-4 Assemble Interactively dialog box



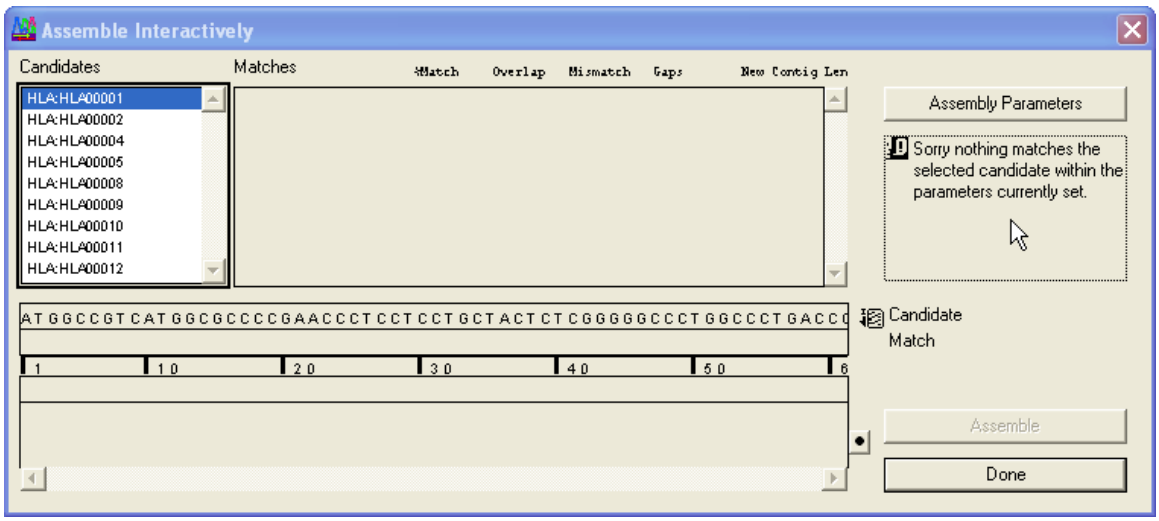
To have Sequencher compare any single sequence or contig with the other items in the **Candidates** list, click on the name of that sequence or the contig in the **Candidates** list. When the comparison is complete, the possible matches appear in the **Matches** list to the right of the **Candidates** list, as shown in Figure 9-5.

Figure 9-5 Interactive assembly showing a candidate and its matches



If Sequencher fails to find a match, as in Figure 9-6, the agent panel displays a message to that effect.

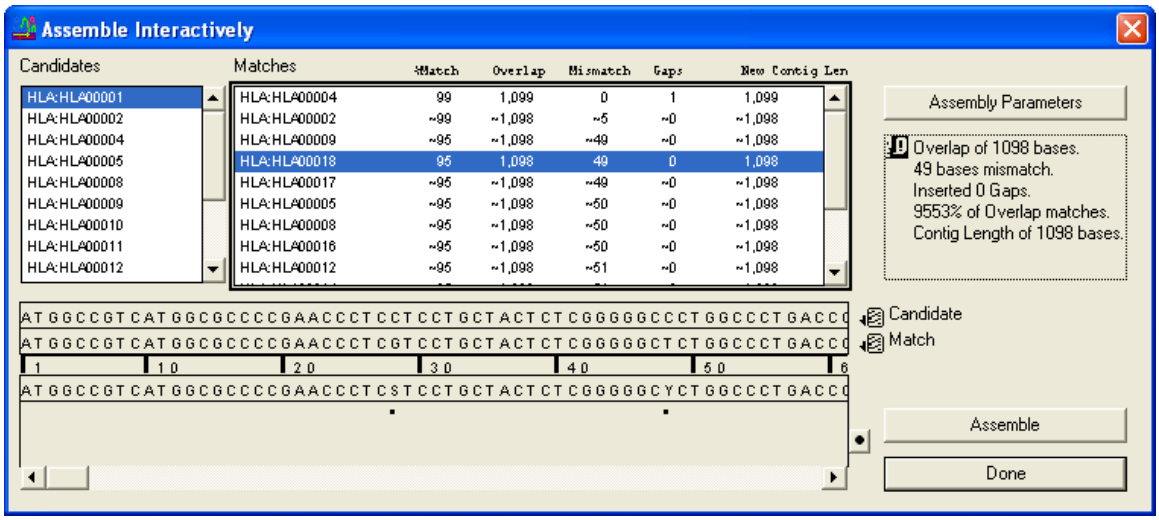
Figure 9-6 Match not found



When you click a possible match in the **Matches** list, Sequencher recalculates the optimal positioning for gaps. The agent then displays a message about the selected sequences and their computed overlap.

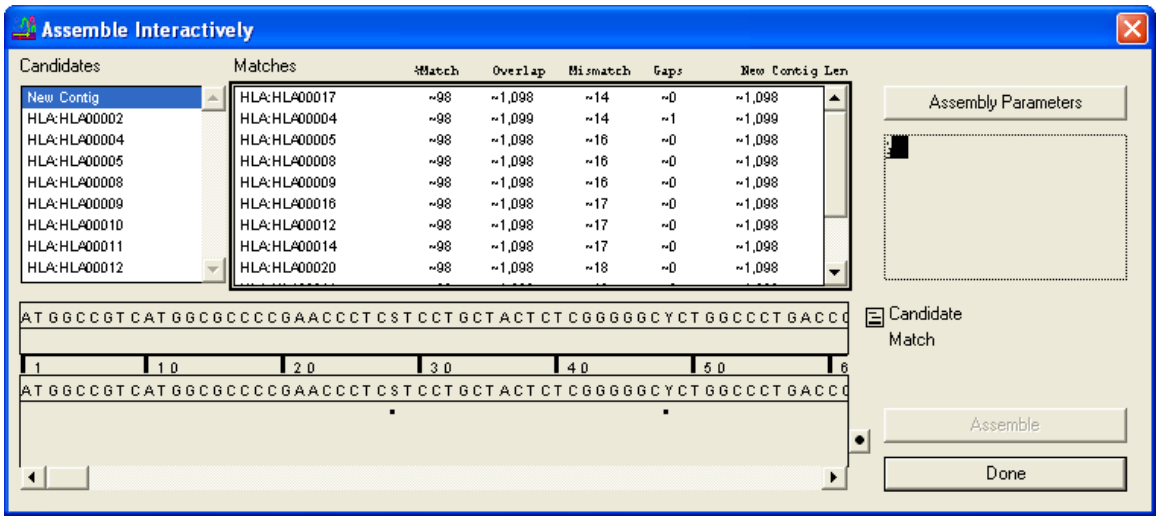
The start of an overlap and the base consensus appear in the fields at the bottom of the interactive assembly box (Figure 9-7). You can explore the alignment further by moving the scroll bar to the right.

Figure 9-7 Interactive Assembly window showing overlap



If you choose to create this assembly, click on the **Assemble** button. Name the new contig in the **Set Name** dialog window and click **OK**. Sequencher takes only a short time to assemble (or reassemble) a contig and the name of the new contig now appears at the top of the **Candidates** list. Note that as a sequence is assembled, its name is removed from the **Candidates** list.

Figure 9-8 The new contig displayed in the Assemble Interactively Window



The new contig will appear in the Project Window when you close the **Assemble Interactively** dialog by clicking on the **Done** button.

Changing parameters for an interactive assembly

If you suspect that your fragments should form a contig but Sequencher cannot suggest candidate matches, you can change the parameters for assembly. To do this while using **Assemble Interactively**, click the **Assembly Parameters** button. The **Assembly Parameters** dialog box will appear. When you have changed the settings to your satisfaction, click **OK** to return to the **Assemble Interactively** window.

Mindlessly Join

Mindlessly Join allows you to put sequences together without using the conventional sets of algorithms. Examples of situations where you might wish to do this include making a plasmid or constructing an artificial cDNA sequence.

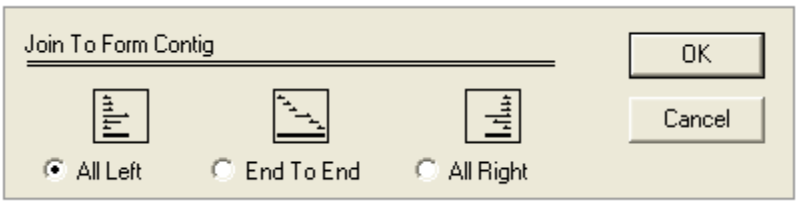
Select the sequences you want to join by dragging a box around them or **Shift**-clicking them.

Note: They will be listed in the new contig in the order in which you selected them.

Go to the **Contig** menu and choose **Assemble Contigs** and then select **Mindlessly Join** from the submenu. If **Assemble by Name** is enabled, the **Mindlessly Join** submenu item becomes **Join by Name**. **Join by Name** uses handles to group the sequences appropriately.

Sequencher will then ask if you want all the sequences aligned at the left of the contig (**All Left**), at the right of the contig (**All Right**) or joined **End to End**. Click the appropriate button. The selected items will be assembled and replaced by a contig icon in the Project Window.

Figure 9-9 Join by Name dialog box



10. Assemble by Name

Many labs apply strict sample-naming conventions to their sequences, with the name indicating something about the data. **Assemble by Name** harnesses the information from your sequence names to automate the process of selecting and naming your contigs.

In this chapter, you will learn how to use **Assemble by Name**, a powerful tool that lets you separately assemble and automatically name multiple contigs from a single selection and assembly command. For example, in just a few clicks of the mouse, **Assemble by Name** will create ninety-six appropriately named contigs from ninety-six forward and reverse sequences.

The Assemble by Name Strategy

The names of your samples frequently contain information such as your primer, template, clone name or sample source. Sequencher provides you with tools to separate the descriptive parts of your sequence name and use these to manage the assembly. These tools are the **Assembly Handles** and the **Name Delimiters**.

An **Assembly Handle** is any set of characters from your sequence name that provides information about that sequence, such as the primer or clone name. A **Name Delimiter** separates each **Assembly Handle** from the following one. Sequence names frequently have several **Assembly Handles** and **Name Delimiters**.

In its simplest form, a **Name Delimiter** may be a character such as a dash or an underscore.

When you use a Name Delimiter, it will look like this:

handle1-handle2-handle3

or

handle1_handle2_handle3.

However, a **Name Delimiter** can also be a combination of letters, numbers and characters.

To define your **Name Delimiter**, Sequencher provides a set of commonly used characters as a drop down menu in the **Assemble by Name Settings** dialog. You will see this menu if you go to **Assembly Parameters** and click on the **Name Settings…** button. When your naming convention does not use simple characters as **Name Delimiters**, you can use formal descriptors known as "regular expressions" instead, which we describe later in this chapter.

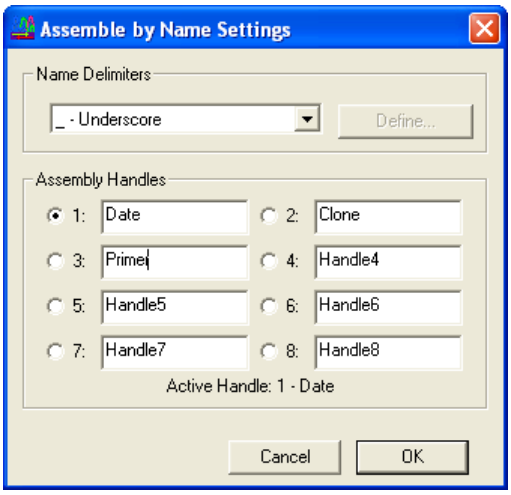
Once you have defined an **Assembly Handle**, you will be able to assemble all of the sequences in your project. Only sequences that share the same **Assembly Handle** will be analyzed for assembly into the same contig. In a clinical application, for example, several contigs might be created with each contig representing sequence data from only a single patient, provided that each sample (patient) has a unique **Assembly Handle** in its name.

Setting the Assemble by Name Conditions

Configuring Assembly Handles using a Single Delimiter

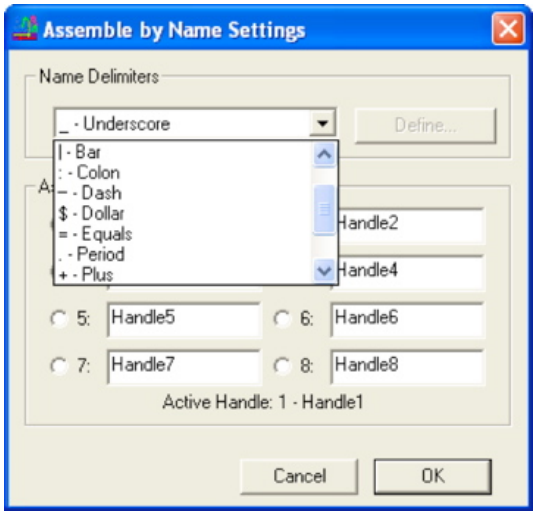
Starting in the Project Window, click on the **Assembly Parameters** button. At the bottom of the **Assembly Parameters** window, click on the **Name Settings…** button. The **Assemble by Name Settings** window appears in Figure 10-1 below.

Figure 10-1 Assemble by Names Settings window



If your **Assembly Handles** are separated by a single character included in the **Name Delimiters** drop down menu, choose that character from the list. Otherwise, follow the directions for an Advanced Expression.

Figure 10-2 The Name Delimiters drop down menu



Configuring Assembly Handles using Advanced Expressions

You will need to create an **Assembly Handle** using a regular expression if your **Assembly Handle** is not separated by one of the characters in the **Name Delimiters** drop down menu.

What is a regular expression?

A regular expression is a way of describing a text pattern using letters, numbers, and special characters. These expressions follow certain syntactical rules.

Starting in the Project Window, click on the **Assembly Parameters** button. At the bottom of the **Assembly Parameters** window, click on the **Name Settings…** button. The **Assemble by Name Settings** window appears.

Choose **Advanced Expression…** from the **Name Delimiters** drop down menu. You will notice that the **Define…** button is now enabled. Click on the **Define…** button.

The **Advanced Expression for Name Parsing** window appears. At the top of the window is a text box where you can type in a regular expression.

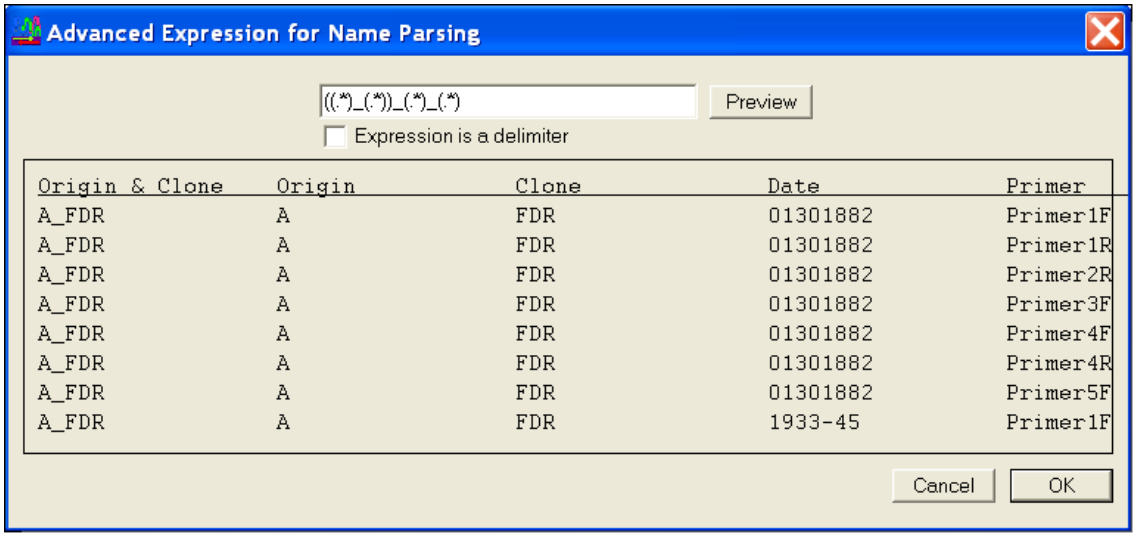
Assemble by Name uses regular expressions in two ways. When the **Expression is a delimiter** box is checked, the regular expression you type in lets you define a delimiter option other than one available from the drop down menu.

If the **Expression is a delimiter** box is *not* checked, your regular expression must completely define each **Assembly Handle** and **Name Delimiter** in your sequence name. The regular expression you write will not work unless it describes the entire name of your sequences. When you type the regular expression into the text field, it should be in the form:

(Assembly Handle1)Name Delimiter(Assembly Handle2)

In the example in Figure 10-3 below, a regular expression is used to combine the first two parts of the sequence name, Origin and Clone, to create a new Assembly Handle, Origin&Clone. The **Preview** button displays the results of your regular expression. If you are satisfied, click on the **OK** button.

Figure 10-3 Advanced Expression for Name Parsing dialog window


Advanced Expression for Name Parsing
((.*)_(.*))_(.*)_(.*)
Preview
Expression is a delimiter
Origin & Clone Origin Clone Date Primer
A_FDR A FDR 01301882 Primer1F
A_FDR A FDR 01301882 Primer1R
A_FDR A FDR 01301882 Primer2R
A_FDR A FDR 01301882 Primer3F
A_FDR A FDR 01301882 Primer4F
A_FDR A FDR 01301882 Primer4R
A_FDR A FDR 01301882 Primer5F
A_FDR A FDR 1933-45 Primer1F
Cancel
OK

For more details on regular expressions, see Appendix 24 “Advanced Expression” and the tutorials in the Tutorials folder.

Setting your Assembly Handle Names

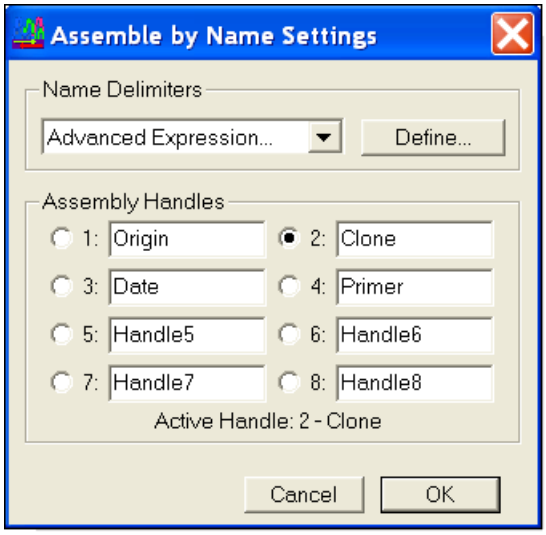
Starting in the Project Window, click on the **Assembly Parameters** button. At the bottom of the **Assembly Parameters** window, click on the **Name Settings…** button. The **Assemble by Name Settings** window appears as in Figure 10-4.

You can describe up to eight **Assembly Handles** of which only one is active at any one time.

For each **Assembly Handle**, you can use the default name Sequencher automatically gives your contigs or you can type a descriptive title into the text box. Click on the radio

button to the left of the handle you wish to activate. Sequencher displays the number and the name of the **Active Handle** at the bottom of this dialog. Click on the **OK** button to dismiss this window and return to the **Assembly Parameters** window.

Figure 10-4 The Assemble by Names Settings window

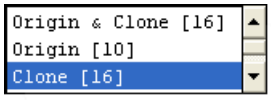


Choosing a New Assembly Handle

There are two ways to choose an **Assembly Handle**. If you are in the **Assembly Parameters** window, you can change the handle by choosing a new one from the drop down menu.

Note: The probable number of contigs that can be formed using this handle is indicated in square brackets to the right of the handle.

Figure 10-5 Assembly Handles drop down menu

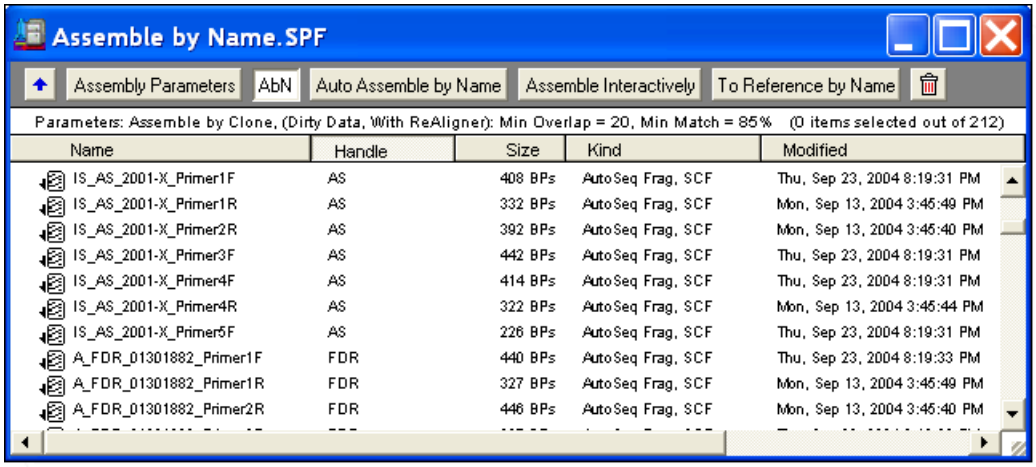


You can also click on the **Name Settings…** button and then click on the radio button next to your preferred handle.

Enabling Assemble by Name

There are two places from which you can enable **Assemble by Name**. From the Project Window, click on the **AbN** button in the button bar. If you are in the List View, you will notice a new column titled **Handle**, which lists the active handle for each sequence. You will also see that the text of the assembly command buttons has changed to **Auto Assemble by Name** and **To Reference by Name**. The new assembly status is also reflected in the current parameters above the column headers.

Figure 10-6 The Project Window with Assemble by Name enabled


Assemble by Name.SPF
Assembly Parameters
AbN
Auto Assemble by Name
Assemble Interactively
To Reference by Name
Parameters: Assemble by Clone, (Dirty Data, With ReAligner): Min Overlap = 20, Min Match = 85% (0 items selected out of 212)
Name
Handle
Size
Kind
Modified
IS_AS_2001-X_Primer1F AS 408 BPs AutoSeq Frag, SCF Thu, Sep 23, 2004 8:19:31 PM
IS_AS_2001-X_Primer1R AS 332 BPs AutoSeq Frag, SCF Mon, Sep 13, 2004 3:45:49 PM
IS_AS_2001-X_Primer2R AS 392 BPs AutoSeq Frag, SCF Mon, Sep 13, 2004 3:45:40 PM
IS_AS_2001-X_Primer3F AS 442 BPs AutoSeq Frag, SCF Thu, Sep 23, 2004 8:19:31 PM
IS_AS_2001-X_Primer4F AS 414 BPs AutoSeq Frag, SCF Thu, Sep 23, 2004 8:19:31 PM
IS_AS_2001-X_Primer4R AS 322 BPs AutoSeq Frag, SCF Mon, Sep 13, 2004 3:45:44 PM
IS_AS_2001-X_Primer5F AS 226 BPs AutoSeq Frag, SCF Thu, Sep 23, 2004 8:19:31 PM
A_FDR_01301882_Primer1F FDR 440 BPs AutoSeq Frag, SCF Thu, Sep 23, 2004 8:19:33 PM
A_FDR_01301882_Primer1R FDR 327 BPs AutoSeq Frag, SCF Mon, Sep 13, 2004 3:45:49 PM
A_FDR_01301882_Primer2R FDR 446 BPs AutoSeq Frag, SCF Mon, Sep 13, 2004 3:45:40 PM

You can enable **Assemble By Name** by clicking on the **Assembly Parameters** button and checking the **Enabled** checkbox at the bottom of the window.

Once you have clicked on the **AbN** button, this box will already be checked.

Setting Assembly Parameters

Sequencher uses the values set in the **Assembly Parameters** window to control the way it assembles sequences. Once you have set your required options, you can store these as your default **Assembly Parameters**. Click on the **Set as Defaults** button at the bottom of the window, and Sequencher will use your preferred **Assembly Parameters** whenever it launches. If you just click **OK** to dismiss the **Assembly Parameters** window, Sequencher will use your new parameters only with this project or session of Sequencher.

Performing an Assembly with Assemble by Name

Once you have chosen your **Name Delimiters** and **Assembly Handles** you are nearly ready to perform your assembly. In addition to the standard **Auto Assemble by Name** command, you can use the **Assemble by Name** function to assemble your sequences with a Reference Sequence or to Mindlessly Join your sequences.

Under the **Contig** menu, the following **Assemble Contigs** menu items will change if you have enabled **Assemble by Name**. The **Mindlessly Join** command becomes **Join by Name…**. Similarly, the **Assemble to Reference** button will now be labeled **To Reference by Name**.

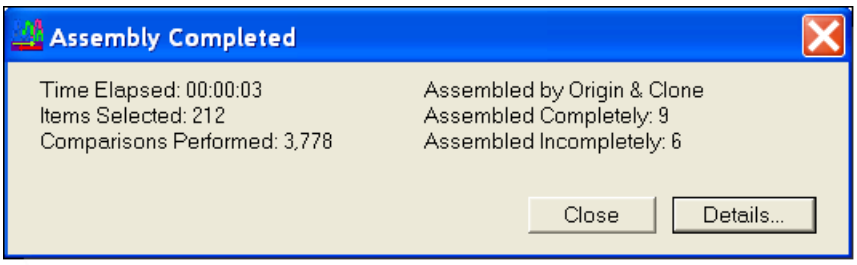
Before you use the **To Reference by Name** command, you will need to define a Reference Sequence. You will also need to choose and set your assembly algorithm and any associated parameters. (For more details, see Chapter 9 “Sequence assembly”.)

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Starting from the Project Window, go to the **Select** menu and click on the **Select All** command to select all the sequences in your project. Click on the **Auto Assemble by Name** button, or the **To Reference by Name** button if you have a Reference Sequence. The **Assembly Preview** window will appear. This window contains the names of the contigs that may be generated using your chosen handle and the probable number of fragments in each contig. Click on the **Assemble** button to continue with the assembly, or click on the **Cancel** button.

Sequencher displays the **Assembly Completed** dialog window when the assembly has finished assembling your sequences.

Figure 10-7 The Assembly Completed window



This dialog window gives a summary of the assembly process. If this information is sufficient, click the **Close** button. This will dismiss the **Assembly Completed** window. If you require more detailed information, click on the **Details…** button. You will see a report that contains further information about the assembly. You can save a text copy of this report by clicking on the **Save As** button. Dismiss the dialog window by clicking the **Close** button.

Note: The name of each contig formed using **Assemble by Name** will be based on the **Assembly Handle** used.

11. The Contig Editor

In this chapter you will learn about the overview and bases view of the contig editor. You will learn how to use these options to explore your contig. We will discuss the steps to take before you start editing: base numbering, setting a consensus calculation and assigning a Reference Sequence.

Contig editing

DNA sequencing is an error-prone process, and so the base calls from an automated DNA sequencer may not correctly represent the sample being sequenced. This means that when working with automated DNA sequencing data it is usually necessary to check and edit your contigs.

Sequencher provides a powerful interface between the user and the data so that you can analyze and edit sequences and contigs according to your own specifications. You can move directly from one ambiguity or disagreement to the next to analyze and override base calling errors, eliminating discrepancies and ambiguities. Sequencher allows you to choose from a number of consensus calculations and continually recalculates the consensus of the contig as you edit the sequences.

Overview of the contig

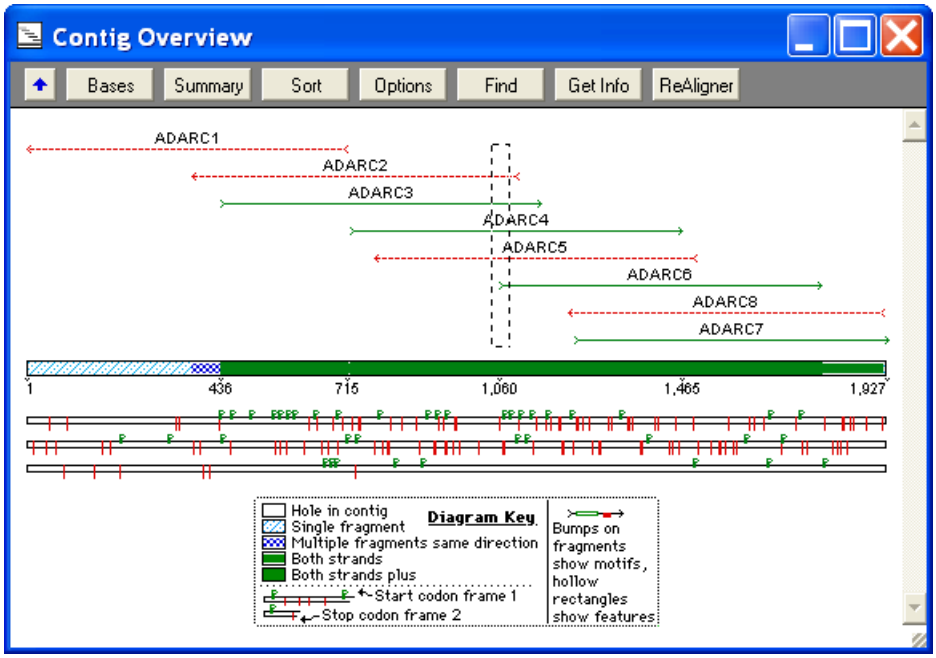
The contig overview window

When you first open a contig by double-clicking on it, you will see an overview that shows how the sequences have been assembled (Figure 11-1). There is also a button bar at the top of the overview window.

The schematic provides you with a wealth of useful information enabling you to assess and manage your data. A horizontal line represents each fragment in the contig. The line will be green and solid if the fragment is in the forward orientation and red and dotted if

it is in the reverse orientation. These orientations may change as more data is added to the contig, unless you have used a Reference Sequence in your assembly. In this case, the Reference Sequence determines the contig orientation. Below the fragments you will see the coverage map and below the coverage map you will see the start and stop map. Notice the diagram key at the base of the window.

Figure 11-1 Contig overview



Selection marquee

In the **Overview** you will notice a dashed rectangle. This animated marquee (sometimes called the marching ants) highlights the section of the contig displayed when you switch to the **Bases View**. You can change the location you want to view by dragging the marquee to a new location. As you move the cursor over the marquee you will see it change to a hand icon. You can now hold down the mouse button and drag the marquee to the desired location.

Checking coverage in the Overview

The bar across the bottom of the overview shows various levels of sequence coverage indicated by different graphic patterns. For instance, a wide solid green line indicates that the sequence coverage comprises both strands. A region with no color or pattern indicates a hole in the coverage. The full array of patterns indicating coverage are shown at the bottom of the Overview in the figure above.

In the example in the figure above, the overview shows the contig starting out with a single, unsupported sequence. Then a second fragment starts to overlap with the first. Next there is an area with "full" coverage: both strands are sequenced with additional confirming fragments. At the extreme right end of the contig there are just two fragments in opposite directions.

Overview features

Bases

To get from the **Overview** to the view used for editing, click the **Bases** button. You can also go to the **View** menu and choose **Bases** from the **Bases, Map, Overview…**submenu. Sequencher displays the contig editing window.

Summary

You can press the **Summary** button to view a compact report of the contig. You can also see this report by going to the **View** menu then choose the **Bases, Map, Overview …** submenu and click the **Summary Report** option. The features of the Summary Report will be dealt with in more detail later (see Chapter 12 “Editing Contigs” for more information).

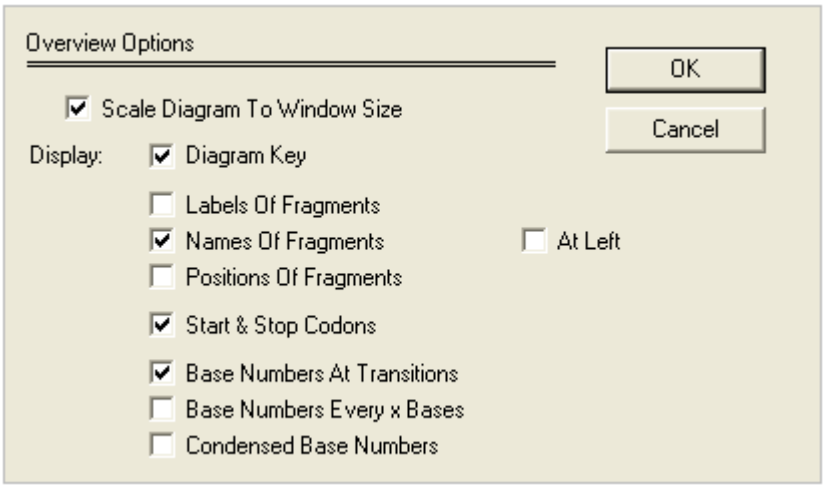
Sort

You can sort the sequences in a contig vertically. Click the **Sort** button to display a dialog window of options. Choose an option by clicking on its radio button and then pressing **OK.** You may also use the sorting options in the **View** menu under the **Sort/Cleanup** submenu.

Overview options

To choose the display options, click the **Options** button or go to the **View** menu and click on **View Options**. An **Overview Options** dialog box will appear where you can set your choices.

Figure 11-2 Contig overview options dialog box



Scale Diagram To Window Size

Selecting **Scale Diagram To Window Size** shrinks the whole overview to fit in the window, even if the names of the individual sequences become unreadable. When you turn this option off, you make the names of the sequences readable, but you may have to scroll the window to see all the sequences in the diagram.

Diagram Key

Clicking this checkbox displays or hides a key that explains the **Overview**.

Labels Of Fragments

You must have already defined and applied some labels to your sequences with **User Preferences** (see Chapter 21, "Customizing Sequencher and User Preferences") and the **Label** command. Clicking **Labels of Fragments** will display or hide any fragment label names.

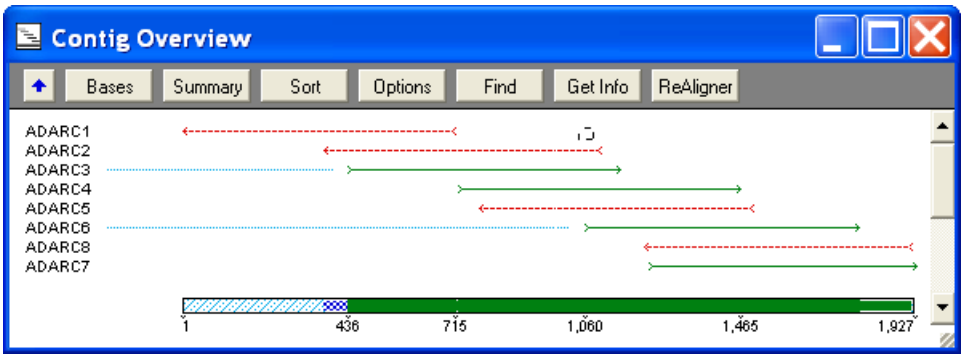
Names Of Fragments

Clicking this checkbox displays or hides the sequence names in the **Overview.** You may have this and the previous item checked at the same time.

Names Of Fragments, At Left

Clicking the **At Left** box in conjunction with **Names of Fragments** lists the fragment names down the left side of the window instead of on top of each arrow (as shown in Figure 11-3). If, for example, the sequences are also sorted by position, a faint blue line will connect the name with its fragment arrow.

Figure 11-3 Names of fragments at left



Positions Of Fragments

Clicking **Positions of Fragments** displays the numerical range positions of the sequences in the **Overview**. If you do not wish to see this information, leave the box unchecked.

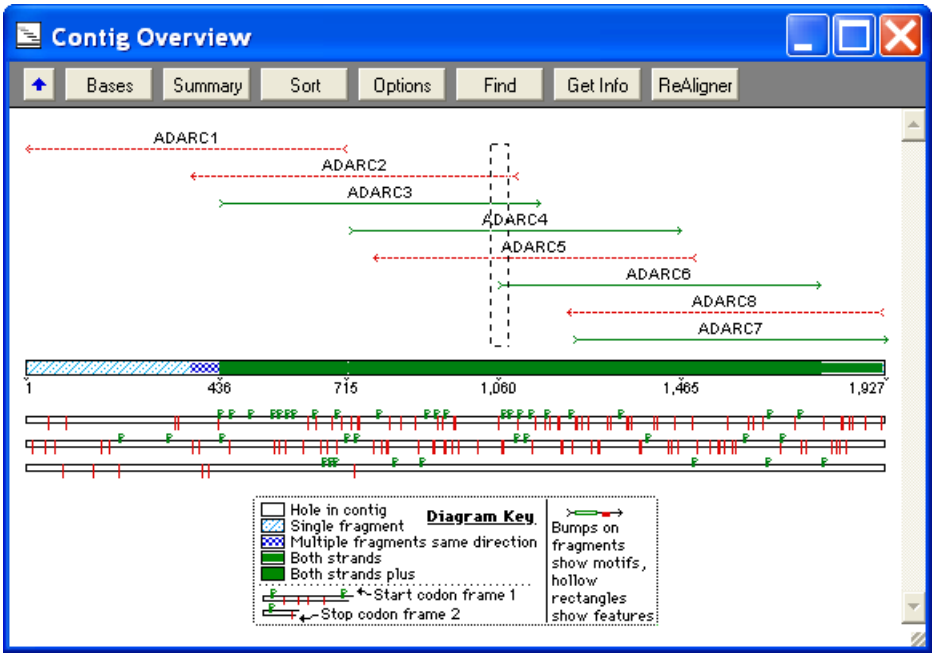
Start & Stop Codons

You can display a start and stop codon map showing all the start and stop codons in the three reading frames of one strand. Start codons are indicated by green flags and stop codons are denoted by vertical red lines. To display this map check the **Start & Stop Codons** box. The map will be positioned at the bottom in the **Overview**.

In Figure 11-4 the three-frame map of start and stop codons is located just above the diagram key.

To see start and stop codons in the other strand, go the **View** menu and choose **Reverse & Comp**. This will reverse and complement your fragments. Choosing **Reverse & Comp** again will restore your contigs original orientation.

Figure 11-4 Start and stop codon map



SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Base Numbers At Transitions

Clicking **Base Numbers at Transitions** displays or hides the numbers showing approximate positions of the coverage transitions.

Base Numbers At Every x Bases

Selecting **Base Numbers Every x Bases** shows the base numbers at constant intervals.

Condensed Base Numbers

Selecting the **Condensed Base Numbers** option changes the font of the base numbers to a compressed version. This option can be used with either **Base Numbers at Transitions** or **Base Numbers at Every x Bases.**

Getting more information

Find

You can search for a specific subsequence in your contig. Fragments, which contain the subsequence, will appear with a heavier line in the **Overview**.

You can choose to find the first occurrence by clicking the **Find** button. You can search for subsequent occurrences using the **Find Next** button. Three radio buttons allow you to control the specificity of the search by choosing **Exact Matches, Specified Bases** (which would allow you to use ambiguity codes such as W to find T or A).

You can use the **Any Ambiguous Base** command, which will find any degenerate DNA that may match. There is also a drop down menu from which you may choose restriction enzyme sequences.

If you have used a Reference Sequence and want to include it in the search, select the **Include Reference Sequence in Find** checkbox

Close the dialog box by pressing **Done.**

Get Info

Summary information about the open contig can be obtained by pressing the **Get Info** button or go to the **File** menu and click on the **Get Info** command.

The Bases View

Working in the Bases View

To get from the **Overview** to the view used for editing a contig, click the **Bases** button. Alternatively, you can go to the **View** menu and then go to the **Bases, Map, Overview…** submenu and choose **Bases**. Sequencher displays the contig editing window shown in Figure 11-5.

Figure 11-5 The contig editing window

To return to the overview click the **Overview** button or go to the **Bases, Map, Overview…** submenu and choose **Overview**.

The labels in Figure 11-6 identify the different parts of the contig editor.

Figure 11-6 Identifying parts of the contig editor

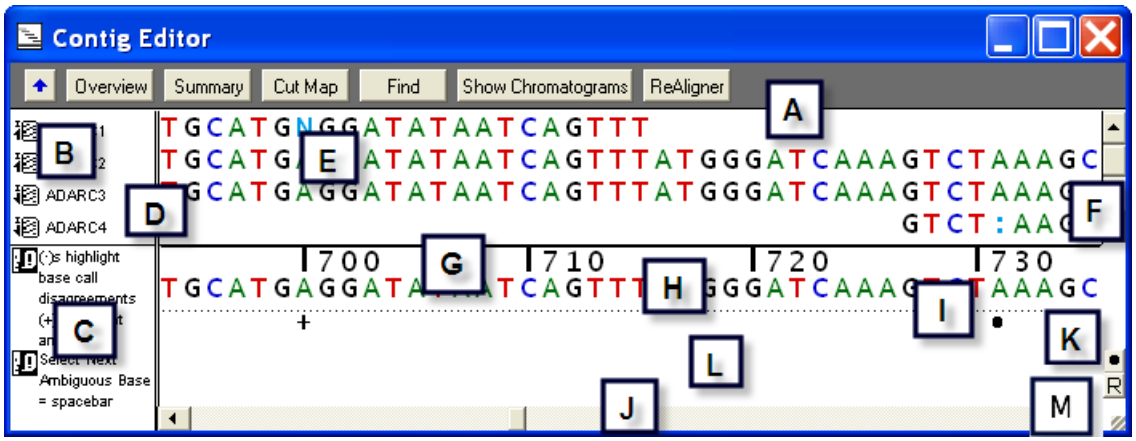


Table 11-1 Key to the Contig Editor

Label	Name	Function
A	**Button Bar**	Contains several buttons you click to perform frequently-used functions.
B	**Sequence list**	Shows which sequences make up the contig you are editing.
C	**Agent**	Describes current selection and the current status of the spacebar shortcut.
D	**Splitter bar**	Controls the width of list/message area and the editing area; dragging the splitter bar to the left or right changes the size

Label	Name	Function
		of these items.
E	**Sequence Bases**	Displays the sequence bases aligned.
F	**Sequence Scroll**	Scrolls up and down through the sequence bases in the contig.
G	**Base Numbers**	Shows which bases you are viewing, numbered from the beginning of the contig.
H	**Consensus line**	Shows the calculated consensus of the sequences that were aligned to form this contig.
I	**Ambiguities**	Displays a plus (+) below any ambiguous base calls in the contig line and a bullet (•) below any base calls with disagreeing bases above them. This area also displays protein translations with a bullet (•).
J	**Base Scroll**	Scroll left and right through the bases of the contig.
K	**Translation Button**	This button toggles the display of protein translations on and off. Five options are available: Frames 1,2, or 3, all three reading frames, or In Frame with the Reference Sequence.
L	**Notation Line**	Displays translation of Reference Sequence and user text notes
M	**Reference Translation Toggle**	This button toggles the display of the Reference Sequence translation on and off.

Viewing sequence names

If the names of your sequences are very long, you may want to move the splitter (item D in the key) to the contig window.

Before you start to edit

Setting the base numbering

You have control over the base numbering of the consensus. For instance, you can select any base in the consensus and make it the "origin" base (base #1). *All the bases before that position will have negative numbers*. To set the numbering, select a base, then go to the **Sequence** menu and choose **Set Base Number** with an appropriate suboption.

Note: If you have already used a Reference Sequence in your assembly, this will set the base numbering.

Consensus calculation

Under the Contig menu there are three choices for calculating the consensus of your sequences: **Consensus Inclusively, Consensus By Plurality,** and **Consensus to Forensic Standards**. The consensus is displayed in the bottom line of the contig editor

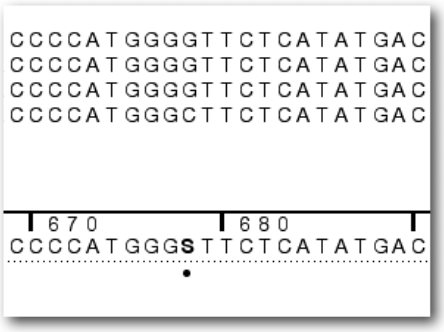
and is continuously updated as you edit. Thus the consensus will alter after *any* change you make, whether that is altering bases or adding new fragments to your contig. Bullets under the consensus highlight discrepancies and pluses indicate ambiguities.

If you choose a consensus calculation for a project and then save and close the project, the consensus choice will be set as the default calculation for that project.

Consensus Inclusively

With **Consensus Inclusively**, Sequencher determines the base in the consensus line by the smallest category of ambiguity that covers all available data. In the example in Figure 11-7, a column with one sequence containing a G and the other containing a C is shown in the consensus line as an S (the IUPAC code for C or G). This consensus is particularly useful in any work involving mutations.

Figure 11-7 Consensus Inclusively example

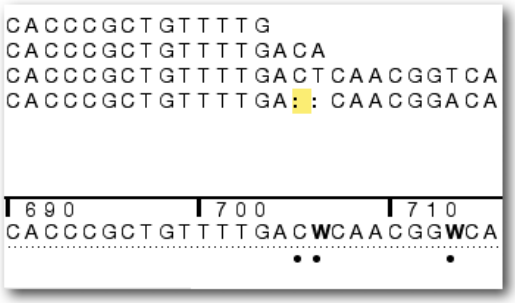


Consensus By Plurality

Consensus By Plurality is Sequencher's default setting. With this consensus calculation, Sequencher determines the consensus line bases by majority rule. If there is no clear majority, Sequencher chooses the smallest category that fits the available data.

In the example in Figure 11-8 below, a column with two sequences that contains a T and an A is still shown in the consensus line as a W, because there is no clear majority. However, a column with three sequences that contains a *gap* in one sequence is shown in the consensus line as an ambiguity code in the consensus line only if the other two sequences do not agree.

Figure 11-8 Consensus by Plurality example

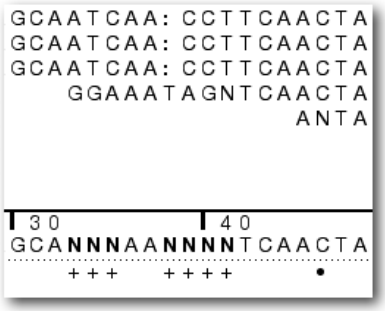


Consensus to Forensic Standards

In certain applications, such as forensic mtDNA sequencing, there is an established standard reference for human identification. This calculation is based on the coverage of fragments at a given position and taken from work at the US Armed Forces DNA Identification Laboratory (AFDIL) in Rockville, Maryland.

- All positions with a coverage of only one fragment are marked as N (ambiguous).
- All IUPAC codes apart from ACG and T are treated as N.
- If any positions disagree where the coverage is between two and four, this position is marked as N. If the coverage is between five and seven, any position with a disagreement will be marked as ambiguous.
- If there are two or more disagreeing bases at a position, this will be called an N regardless of the coverage.

Figure 11-9 Consensus to Forensic Standards example



Assigning a Reference Sequence

To designate a sequence as a Reference Sequence, select the sequence icon in the **Project Window** and check the **Reference Sequence** item in the **Sequence** menu. From now on, that sequence icon will include a small letter "R" to remind you that it has been marked as a Reference Sequence (see Chapter 7 “The Reference Sequence” for more information).

12. Editing Contigs

In this chapter we explain how to find bases such as ambiguities and low-confidence bases. We discuss how to perform your edits, move and delete bases and sequences, insert gaps or bases and create a new sequence from a consensus. You will learn how to compare sequences to highlight differences and how to create customized reports.

Finding bases which need attention

You are likely to see a number of issues in your data. Some will need to be examined and perhaps edited. These issues might be indicated as an N, as a disagreement, as a gap, or as having a low confidence rating. You can use one of the **Next** commands from the **Select** menu in the consensus sequence of a contig to find any of these problem areas.

Once you have used any of the **Next** commands, Sequencher will activate the space bar to execute that operation. If you use either the menu command or the shortcut key combination two consecutive times, Sequencher will remind you that this easier alternative is available.

Viewing an individual sequence from the contig

To look at the data for one of the sequences in your contig, double-click the sample name in the sequence list. Sequencher opens a locked sequence viewer containing the data for that sequence.

You can also select a sequence in the contig editor and then go to the **File** menu and choose **Open Window**.

Finding disagreeing ambiguities

You can find ambiguities quickly by typing **Ctrl+N** or by clicking in the consensus line and going to the **Select** menu and choosing **Next Ambiguous Base**. This will find data that is indicated as an N or as a disagreement or gap. The **Next Ambiguous Base** command will also find positions in the contig where the only contribution to the

consensus is from a base with a low confidence score. This type of base will be marked with a + (plus) or a • (bullet) below the consensus line.

Each time you press **Ctrl+N** Sequencher moves the cursor to the next ambiguity in the consensus, starting *after* the currently selected character. After establishing this function, you can move to the next ambiguous base by hitting the spacebar or going to the **Select** menu and using the **Repeat Select** command. If you use the **Shift** key in conjunction with a **Next Ambiguous Base** key combination, the cursor will move in the reverse direction.

Finding Disagreeing Bases

You can find disagreeing bases quickly by typing **Ctrl+D** or clicking in the contig line, going to the **Select** menu and choosing **Next Contig Disagree**. This type of base will be marked with a • (bullet) below the consensus line

Sequencher moves the cursor to the next disagreement in the consensus, starting *after* the currently selected character and ignoring ambiguities.

Finding Low Confidence Bases

You can quickly locate low confidence bases by typing **Ctrl+L** or by going to the **Select** menu and choosing **Select Next Low Confidence Base**. You can set the ranges for Low, Medium and High confidence in the **User Preferences** (see Chapter 21 "Customizing Sequencher and User Preferences").

Note: You can use the **Next Low Confidence Base** command to help direct your editing. Under the **Window** menu click on **User Preferences…** and set the **Low Confidence** range appropriately for the value range you wish to find. Go to the **Select** menu and click on **Next Low Confidence Base** to move your cursor directly to the bases that require attention.

Finding Edited Bases

You can quickly review edits already made to a contig by typing **Ctrl+E** or by going to the **Select** menu and choosing **Next Edited Base.**

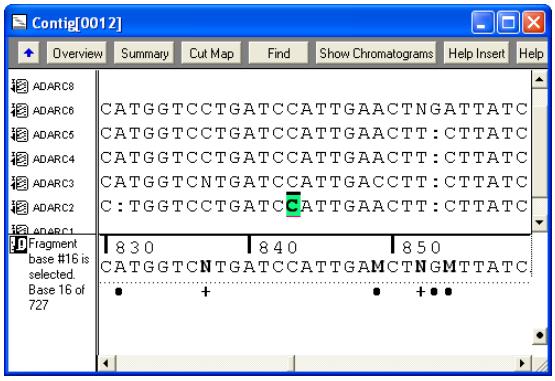
Note: This function will not find deletions.

Performing edits

The Edit Command

To edit a base, first set the **Edit** mode you wish to use, then select the base by clicking on it (see Figure 12-1). Now type the new character over the old. If you accidentally type the wrong character, go to the **Edit** menu and choose **Undo** or just type over the mistake. Sequencher will recalculate the consensus line automatically. If you resolve an ambiguity at a particular position, the ambiguity marker below the consensus line disappears (see Figure 12-1).

Figure 12-1 A base is selected in the contig editor



Editing modes

The **Edit** command allows you to replace/overstrike, insert or erase a base call. The specific option you choose from the **Edit** menu will create that behavior. When you use **Replace When Editing**, the base you have just edited will remain highlighted. When you use **Overstrike When Editing**, the cursor will move one base to the right and this base will then be highlighted. **Insert When Editing** allows you to insert bases before the currently selected base, pushing all of the following bases forward. This also allows you to add bases beyond the end of a sequence. Sequencher uses **Replace When Editing** as its default setting.

Note: If you insert or delete a base from any one sequence, the sequence may no longer align with other sequences after the point where you made your change. The consensus line will reflect this by showing lots of ambiguities in the form of bullets •. If you have made only one change, go to the **Edit** menu and the **Undo** command will remove it.

Viewing Base Edits

To assess the quality of your data, you need to be able to see how much modification has already been done to your contig as you edit it.

When a base is edited, Sequencher changes its appearance by displaying it in magenta and bold-faced text. If you want edits to stand out even more (for instance in a black and white printout), you can also invert the case of the characters (e.g. change all edits from uppercase to lowercase text).

To see the options for displaying edits, go to the **View** menu and click on the **Base Edits As…** submenu to choose from **Not Highlighted**, **Bold Magenta**, or **bOLD and cASE cHANGE**.

Collect Gaps

The **Collect Gaps** command allows you to select a range in a fragment or contig consensus and move all of the gaps within this selection to either the 5’ or the 3’ end of the selection.

If you wish to collect the gaps within a single fragment, you must select the range within that fragment. Go to the **Sequence** menu and choose the **Collect Gaps** command. To

move the gaps to the 5’ end, select **Left** from the submenu. To move gaps to the 3’ end of the selection, choose **Right** from the submenu.

If you wish to collect the gaps within a region of a single contig you must select the range from the consensus line. Go to the **Sequence** menu and choose the **Collect Gaps** command. To move the gaps to the 5’ end, select **Left** from the submenu. To move gaps to the 3’ end of the selection, choose **Right** from the submenu.

Note: The **Collect Gaps** command may have the effect of increasing the number of ambiguities. Check the consensus line to see if the number of ambiguities has increased or decreased. You can go to the **Edit** menu and use the **Undo** command to reverse the **Collect Gaps** command.

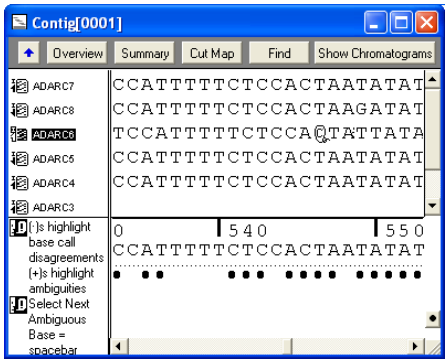
Moving Bases and Gaps

You can move a selection within a single sequence using the lasso tool. The selection can be a single base or gap or a number of bases or gaps.

Begin by making your selection by dragging the cursor across the bases or gaps you wish to move. The region will be highlighted. Then hold down the **Alt** key and your cursor will turn into a little “lasso” tool. Hold down the **Alt** key and click on the highlighted selection you want to move. Then drag it to its new location. The selection will change to **Bold Magenta** text in its new position.

Note If you have used the **Large Gap** algorithm to assemble sequences you can double click anywhere within the gap region to select it.

Figure 12-2 The lasso tool around a base to be moved



Using ReAligner to clean up contigs

The **ReAligner** button on the Contig Editor button bar lets you apply the **ReAligner** algorithm to optimize gap placement in your sequences without first dissolving your contig. This is especially useful when you are left with gaps and unaligned bases after removing one or more sequences from a contig.

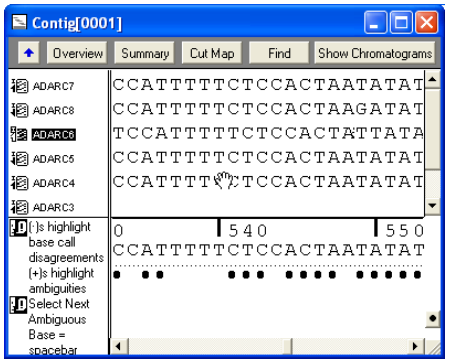
Click the **ReAligner** button. Once the ReAlignment has finished, a new dialog window called **ReAligment Complete** will appear. This window displays a summary of the disagreements, gaps and ambiguities before and after the ReAlignment.

Note: ReAlignment is not the same as reassembly. If you wish to reassemble your sequences you must first dissolve your contig. (See Chapter 9 "Sequence Assembly" for more information.)

Moving sequences

If you hold down the **Ctrl** key, the "grabber" tool is invoked. It resembles a small hand. You can use the grabber tool to drag an entire sequence to the left or right. This method of moving entire sequence fragments also works in the **Overview**. The grabber tool is shown in Figure 12-3. You may notice that some or all of your sequence no longer aligns and that a large number of bullets (•) appear in the consensus line. If you decide that the new position is not correct you can use the **Undo** command to restore it to its previous place.

Figure 12-3 Grabber tool

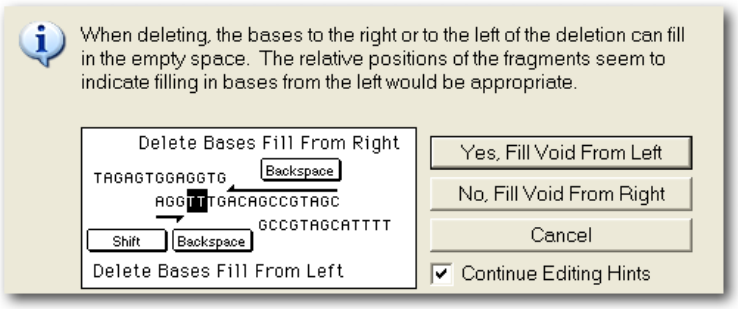


Note: Moving sequences in a low coverage region could leave a hole in the middle of the contig. This is not recommended since the hole may cause the contig to align inappropriately with other data. Sequencher will warn you if this about to happen.

Deleting bases

You can delete bases in the middle of a sequence by simply highlighting them and then using the **Delete** key. Sequencher asks if you want to fill the gap with bases moved from either the left or the right of the void you have created. Click on either **Yes, Fill Void From Left** or **No, Fill Void From Right** button. (Figure 12-4.) Sequencher then moves the rest of the bases accordingly. Click on **Cancel** if you do not want to delete anything. You can also perform the deletion by going to the **Sequence** menu, using the **Delete Bases** command and then choosing either **Fill Void From Left** or **No, Fill Void From Right** from the submenu.

Figure 12-4 Fill Void dialog box


When deleting, the bases to the right or to the left of the deletion can fill in the empty space. The relative positions of the fragments seem to indicate filling in bases from the left would be appropriate.
Delete Bases Fill From Right
TAGAGTGGAGGTG
Backspace
AGGTTTGACAGCCGTAGC
GCCGTAGCATTTT
Shift
Backspace
Delete Bases Fill From Left
Yes, Fill Void From Left
No, Fill Void From Right
Cancel
Continue Editing Hints

Deleting bases from the 5' or 3' ends

Although you may have automatically trimmed your sequences for poor quality data when you imported them (see Chapter 6 "Preparing your data for assembly"), you may still need to delete a few 5' bases to correct any mistakes made during the amplification process.

Select the bases you want to delete. If you are at the 5' (left end) then click the **Forward Delete** key. Sequencher will delete those bases without moving the entire sequence to the left. Similarly, if you choose bases on the 3' (right) end use the **Backspace** key and Sequencher will automatically delete those bases without realigning your sequence.

Note: When editing from the consensus line you do not have to worry about the bases shifting to the left and right.

Inserting gaps or bases into a contig

If you want to insert gaps, go to the **Sequence** menu and click on **Insert Gaps & Move Bases** then choose **Right** or **Left** from the submenu. You can also create gaps by selecting a number of bases equivalent to the gap size you want to create. Press the **Tab** key if you want the selected bases to move to the right. To move to the left, hold down the **Alt** key when you press **Tab.** The space created will be filled with gap marks (small colon marks) that will appear without disrupting the rest of your contig.

Once your gaps are in place you can either type the bases you want into the gap using **Overstrike When Editing** mode or use the lasso and grabber tools to move existing bases into the new position.

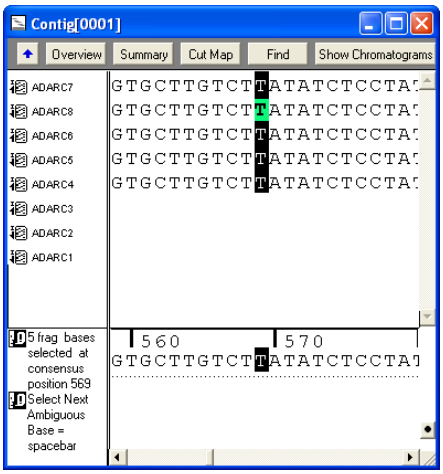
Making edits from the consensus line

You can simultaneously edit all the bases in the column above the position you choose in the consensus line, instead of individually changing each one. To do this, select a base in the consensus line and enter the character that represents the edit you want to make in that column. If you currently have a base selected in a sequence fragment, go to the **Select** menu and choose **Contig Column (Ctrl+K).**

Any bases in the sequences above the contig column you select that do not match the character you type *will be changed* to match the character you type in the consensus line. Your edits will appear according to the preferences you set in the **View** menu. Generally this will be magenta and bold-text.

You can also delete bases from the consensus line. You cannot use the **Undo** command to reverse this action if your contig contains a Reference Sequence.

Figure 12-5 The base is changed



Note: Remember that changing bases from the consensus line edits only the sequences, after which the consensus is recalculated and displayed. Your action does *not* change the consensus directly. Changes in the consensus result from matching sequences with each other and evaluating them according to the type of consensus calculation you have chosen (e.g. by plurality, inclusively, forensic).

Further note that this type of editing does not change Reference Sequences.

Split After Selection…

You can split a sequence fragment into two pieces. Put the cursor where you want to split the sequence. Go to the **Sequence** menu and choose **Split After Selection….** Sequencher will ask you if you want to split the sequence after the last currently selected base. Click on the **Split** button to proceed. Sequencher will then split the sequence in two and label the new fragments for you.

You can also apply the **Split After Selection…** command to an entire contig by simply making your selection in the consensus sequence, and then choosing **Split After Selection…** from the **Sequence** menu.

Auto Match

Sometimes you may wish to edit a number of sequences simultaneously, perhaps to match a high quality segment of sequence. Sequencher allows you to perform this type of mass edit on a range of bases using the **Auto Match** command.

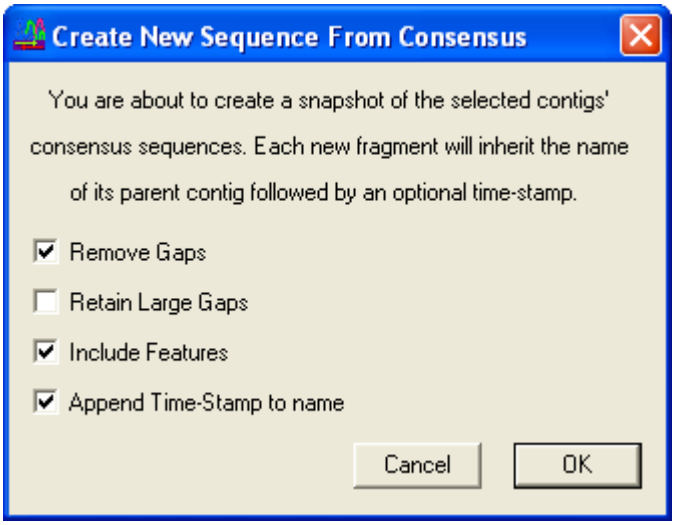
First select the base or range of bases to which the other sequences should match, then go to the **Edit** menu and choose **Auto Match**. This command will edit (and highlight) every overlapping base so that it will agree with the selected sequence. To reverse this procedure go to the **Edit** menu and click on **Undo**.

Create new sequence from a consensus

You can create a new sequence from the consensus sequence by going to the **Contig** menu and choosing **Create New Seq From Consensus...** In addition to creating a new sequence you can use this feature to track the changes made to your contig over time.

Sequencher displays a dialog box that tells you that you are about to create a snapshot consensus of the selected contig. This dialog allows you to control some basic features of the new sequence. Figure 12-6 shows the options dialog box for this command.

Figure 12-6 Create New Sequence From Consensus options dialog box


Create New Sequence From Consensus
You are about to create a snapshot of the selected contigs' consensus sequences. Each new fragment will inherit the name of its parent contig followed by an optional time-stamp.
Remove Gaps
Retain Large Gaps
Include Features
Append Time-Stamp to name
Cancel
OK

Check the box to indicate how you want the new sequence to appear. If the default setting **Remove Gaps** is checked then Sequencher will remove all gaps (large and small) when it creates the new sequence.

If you want to keep large gaps independently from other gaps then select the checkbox called **Retain Large Gaps**. In Sequencher a large gap consists of 10 or more consecutive bases. Such a gap might appear if, for example, you have used the **Large Gap** algorithm to assemble genomic and cDNA. The **Retain Large Gaps** option is not available if **Remove Gaps** is checked.

Note: To see the new sequence go to the Project Window. Remember this is a snapshot of the consensus at the time of its creation. Future edits will not be marked.

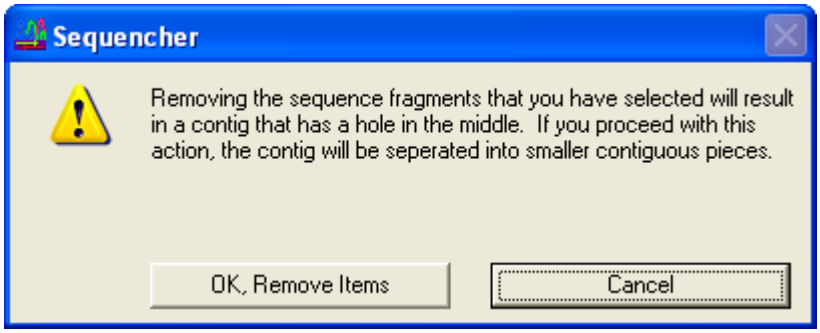
Removing sequences from a contig

To remove sequences from a contig, click on the name of the sequence you wish to remove from the list at the left side of the **Contig Editor** window or in the **Overview** window. To remove more than one sequence at a time, hold down the **Shift** key while you click the names. Next, go to the **Contig** menu and choose **Remove Selected Sequences…** Sequencher will ask you if you are sure you wish to remove the sequence(s). Click on the **Remove Fragments** button to delete the sequences from the contig.

The sequences you just removed from the contig are moved out of the contig and back into the **Project Window**. They will also stay selected.

Note: This action may break a contig into multiple pieces. If this is about to happen, Sequencher will display a dialog box (Figure 12-7) to ask if this is really what you intend to do. Press the **Cancel** button if you do not wish to proceed. Any new contigs generated in this way are named by appending an alphabetic character to the name of the original contig.

Figure 12-7 Warning dialog when about to make a hole in a contig


Sequencher
Removing the sequence fragments that you have selected will result in a contig that has a hole in the middle. If you proceed with this action, the contig will be seperated into smaller contiguous pieces.
OK, Remove Items
Cancel

Dissolving a contig

To dissolve a contig, first select the contig's icon from the project. Go to the **Contig** menu and choose **Dissolve Contig…**. You can also perform this command while you are in the contig editor.

Sequences from the dissolved contig are shown highlighted in the project. It is important to remember that all edits made to a sequence while in the contig will be retained. If you want to restore the original sequences you must go to the **Sequence** menu and click on **Revert to Experimental Data**.

13. The Variance Table

In this chapter we explain how to use the Variance Table, an interactive display of the differences among your sequences or contigs. We go on to discuss how you find differences and work with your trace and confidence data. The Variance Table is useful for SNP analysis, mutation detection, and clone checking.

The Variance Table Strategy

What is a Variance Table

In its simplest form, a Variance Table compares and displays the differences between two sequences. The data in the Variance Table are dynamically linked to the data in the underlying contigs.

You can compare some or all of the sequences in a contig to the **Consensus** sequence, the **Reference Sequence** or the **Top Sequence** of that contig (see Figure 13-1 below). This generates a Variance Table which will display the differences between the exemplar sequence and your chosen sequence(s). The results may be from just one contig, but can represent an assembly of hundreds or even thousands of sequences. If you are performing activities such as de novo sequencing, clone checking or resequencing, you may want to use this form of the Variance Table.

The Variance Table and consensus sequences

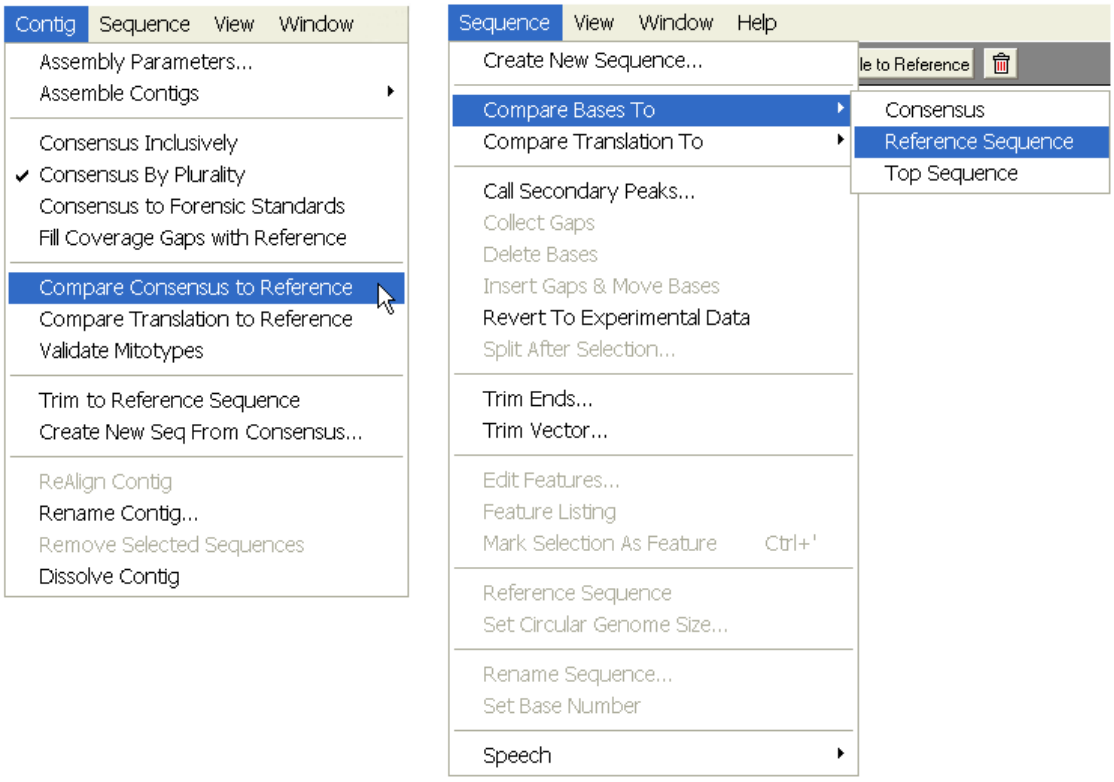
You can also create a Variance Table that summarizes all of the differences between selected contig consensus sequences (see Figure 13-1 below) in your project and a common Reference Sequence (consensus Variance Table). When you are working with multiple samples from multiple sources, and have used **Assemble by Name** with a Reference Sequence, you should use this Variance Table. If you are performing activities such as comparative sequencing, clone checking or SNP analysis, you probably want to use this form of the Variance Table.

The Translated Variance Table

If you are interested in focusing on the differences between sequences or contigs at the amino acid level you should use the Translated Variance Table. This table differs from the Variance Table by displaying both codons and their associated amino acid residues. Each row in the Translated Variance Table summarizes the variations among all of the selected sequence translations at a given amino acid position *relative to the exemplar*.

You can also create a Translated Variance Table that summarizes the differences between the translations of selected contig consensus sequences in your project and the translation of a common Reference Sequence (Translated consensus Variance Table). When you are working with multiple samples from multiple sources, and have used **Assemble by Name** with a Reference Sequence, you should use this form of the Translated Variance Table.

Figure 13-1 Variance Table Menu Commands



The Structure of the Variance Table

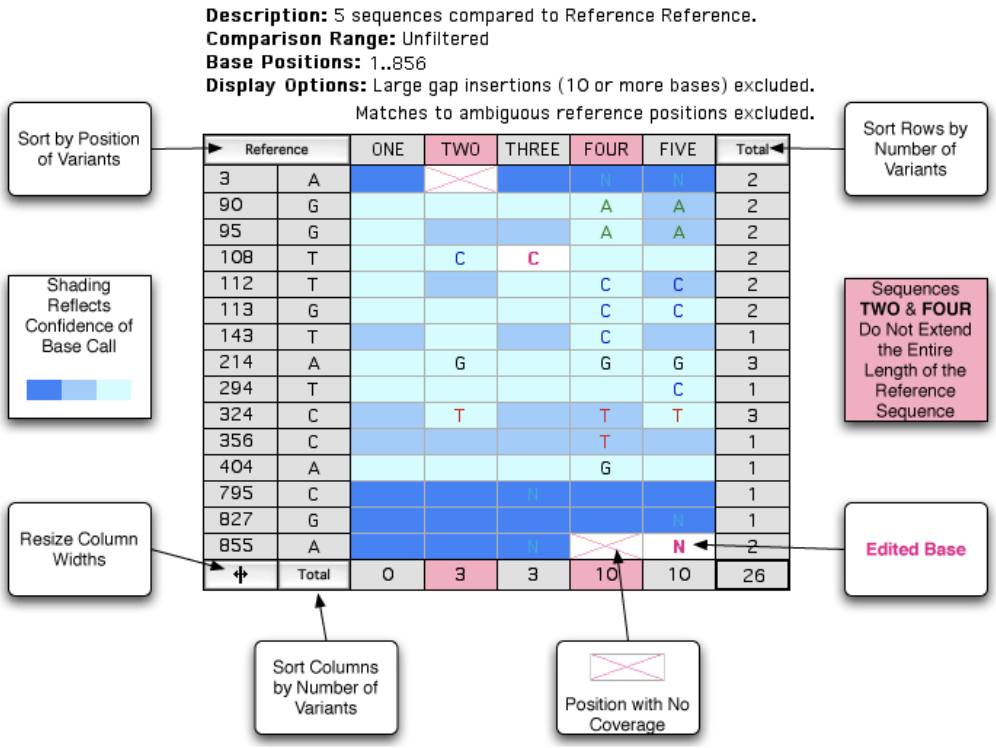
Each row in a Variance Table summarizes the variation among all of the selected sequences at a given base position *relative to the exemplar* (see Figure 13-2). The exemplar can be a Reference Sequence, a consensus sequence, or any other sequence you choose. In the consensus Variance Table the exemplar is limited to the Reference Sequence.

The range of bases included in a Variance Table is called the Comparison Range. The length of the exemplar determines the default Comparison Range and sets the base numbering.

The extent to which the exemplar matches any sample sequence is called the Coverage Range. The Coverage Range can be complete (full range) or incomplete. If the range is full, the entire length of the sample sequence has been compared to the entire length of the exemplar sequence.

If the range is listed as incomplete, the display of the sample sequence comparison runs along only part of the exemplar sequence. If a sequence extends the full length of the Comparison Range, its column header will be shaded in gray. If a sequence in the Table does not extend the full length of the Comparison Range, its column header will be shaded in pink.

Figure 13-2 Components of a Variance Table



The first column of the Variance Table displays the position and the second column displays the base call of the exemplar sequence, but *only at positions that differ*. The additional columns in the Variance Table are created by the sequences or contigs you select for comparison.

You can view or edit the data in any version of the Variance Table by double clicking on any of the cells in that table. In **Review** mode, when you type a new base in the Variance Table, you edit the underlying sequence or contig automatically.

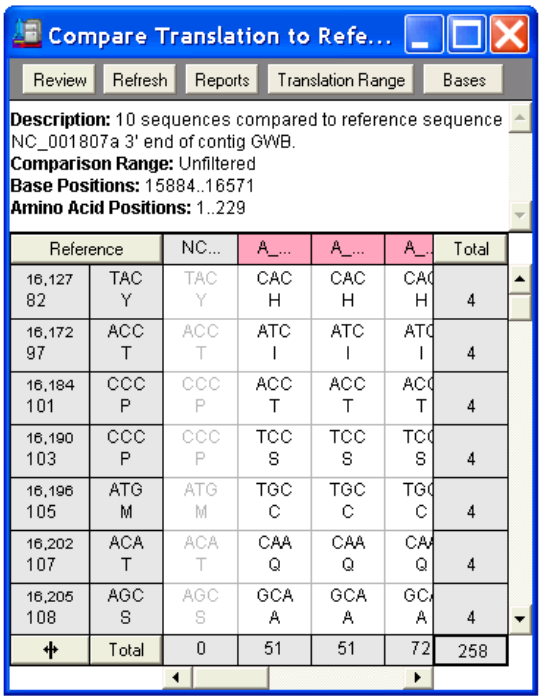
Structure of the Translated Variance Table

The structure of the Translated Variance Table follows the format of the Variance Table but its display consists of an amino acid and its codon in each cell (see Figure 13-3).

When you see a cell in the table with an amino acid and its codon written in black typeface, this indicates a difference between your exemplar and your sample. You may also see some cells in the table where the amino acid and codon are displayed in light gray. Amino acids in these cells match the exemplar but use either an identical or synonymous codon.

The numbers which appear in the first column of the Translated Variance Table refer to the position of the first base of the codon (in the upper location) and the position of the amino acid (in the lower one).

Figure 13-3 The Translated Variance Table


Compare Translation to Refe...
Review
Refresh
Reports
Translation Range
Bases
Description: 10 sequences compared to reference sequence NC_001807a 3' end of contig GWB.
Comparison Range: Unfiltered
Base Positions: 15884..16571
Amino Acid Positions: 1..229
Reference
NC...
A_...
A_...
A_..
Total
16,127 82 TAC Y TAC Y CAC H CAC H CAC H 4
16,172 97 ACC T ACC T ATC I ATC I ATC I 4
16,184 101 CCC P CCC P ACC T ACC T ACC T 4
16,190 103 CCC P CCC P TCC S TCC S TCC S 4
16,196 105 ATG M ATG M TGC C TGC C TGC C 4
16,202 107 ACA T ACA T CAA Q CAA Q CAA Q 4
16,205 108 AGC S AGC S GCA A GCA A GCA A 4
Total 0 51 51 72 258

Note: When Sequencher encounters a gap, it skips to the next available base to include it as a component of the codon. If an N is in the first or second position of the codon, Sequencher displays a question mark (**?**) instead of an amino acid. If the N is in the third position of the codon, a residue may be displayed. This will depend on the genetic code redundancy.

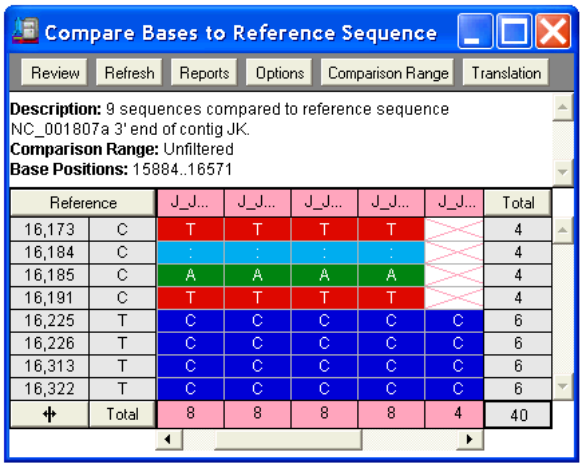
Creating a Variance Table

Variance Table for sequences in the same contig

When you are ready to begin creating your Variance Table, either select individual sequences from within a contig, or go to the **Project Window** and select a contig by clicking on its icon. This will select all of the contig's sequences.

Then go to the **Sequence Menu** and choose **Compare Bases To** and either **Consensus, Reference Sequence** or **Top Sequence** from the submenu. The Variance Table appears in a new window (Figure 13-4).

Figure 13-4 Variance Table Created with Compare Bases To Reference Sequence


Compare Bases to Reference Sequence
Review
Refresh
Reports
Options
Comparison Range
Translation
Description: 9 sequences compared to reference sequence NC_001807a 3' end of contig JK.
Comparison Range: Unfiltered
Base Positions: 15884..16571
Reference
J_J...
J_J...
J_J...
J_J...
J_J...
Total
16,173 C T T T T 4
16,184 C : : : : 4
16,185 C A A A A 4
16,191 C T T T T 4
16,225 T C C C C C 6
16,226 T C C C C C 6
16,313 T C C C C C 6
16,322 T C C C C C 6
Total 8 8 8 8 4 40

The consensus Variance Table

When you want to compare a consensus sequence to a Reference Sequence you use the **Contig** menu's **Compare Consensus to Reference** command.

You can also use the **Compare Consensus to Reference** command to compare the consensus sequences from multiple contigs that share the same Reference Sequence. These contigs are the result of using the **Assemble By Name** command with a Reference Sequence.

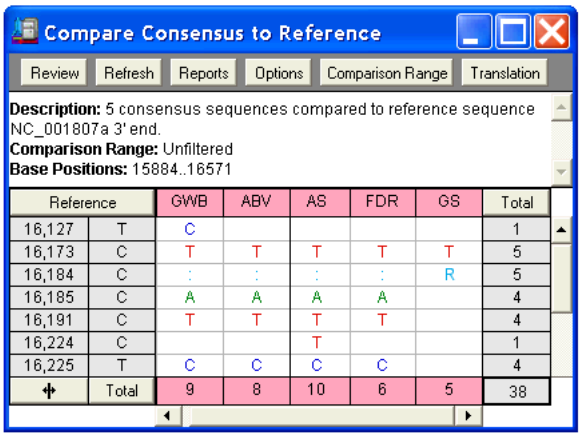
For each of these operations, go the Project Window and select one or more contigs that share a common Reference Sequence. Then, from the **Contig** menu, choose the **Compare Consensus to Reference** command. The Variance Table appears in a new window (see Figure 13-5).

Each cell in the resulting Variance Table will display a consensus base when that base differs from the Reference Sequence. Pink shading in the header and footer cells of the Variance Table highlight contigs where the consensus sequence does not cover the span of the Reference Sequence. The pink X's mark specific cells where coverage is missing

(see Figure 13-5). It is possible for a sequence to have partial coverage, a pink header and footer, but not have any pink X's in the column.

Note: When comparing multiple consensus sequences, the **View** menu's **Display Base Confidences** option is not relevant, because confidence values exist only in fragments and not in consensus sequences.

Figure 13-5 Compare Consensus to Reference Variance Table



For more information about **Assemble by Name** and Reference Sequences, see Chapter 10 "Assemble by Name" and Chapter 7 "The Reference Sequence".

Translated Variance Table for sequences in the same contig

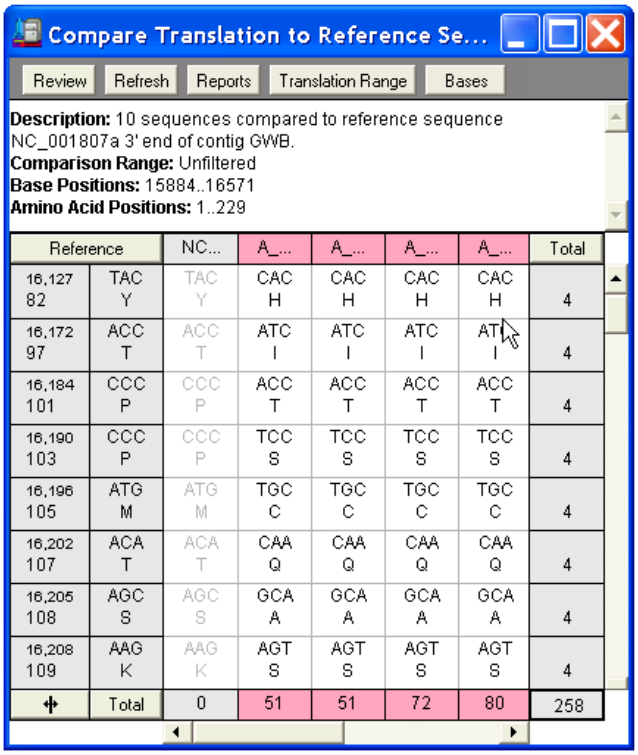
When you are ready to begin creating your Translated Variance Table, either select individual sequences from within a contig or go to the **Project Window** and select a contig by clicking on its icon. This will select all of the contig's sequences.

Then go to the **Sequence** menu and choose **Compare Translation To** and either **Consensus, Reference Sequence** or **Top Sequence** from the submenu. The Translated Variance Table appears in a new window (see Figure 13-6).

The Translated Variance Table header summarizes the table's contents including a **Description**, the **Base Positions** and the **Amino Acid Positions**. If a sequence in the Table does not extend the full length of the Comparison Range, its column header will be shaded in pink. The pink X's mark cells that are not covered by sequence data, indicating that the status of the base at that position is unknown.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Figure 13-6 Translated Variance Table for sequences in the same contig



Reference		NC...	A_...	A_...	A_...	A_...	Total
16,127 82	TAC Y	TAC Y	CAC H	CAC H	CAC H	CAC H	4
16,172 97	ACC T	ACC T	ATC I	ATC I	ATC I	ATC I	4
16,184 101	CCC P	CCC P	ACC T	ACC T	ACC T	ACC T	4
16,190 103	CCC P	CCC P	TCC S	TCC S	TCC S	TCC S	4
16,196 105	ATG M	ATG M	TGC C	TGC C	TGC C	TGC C	4
16,202 107	ACA T	ACA T	CAA Q	CAA Q	CAA Q	CAA Q	4
16,205 108	AGC S	AGC S	GCA A	GCA A	GCA A	GCA A	4
16,208 109	AAG K	AAG K	AGT S	AGT S	AGT S	AGT S	4
	Total	0	51	51	72	80	258

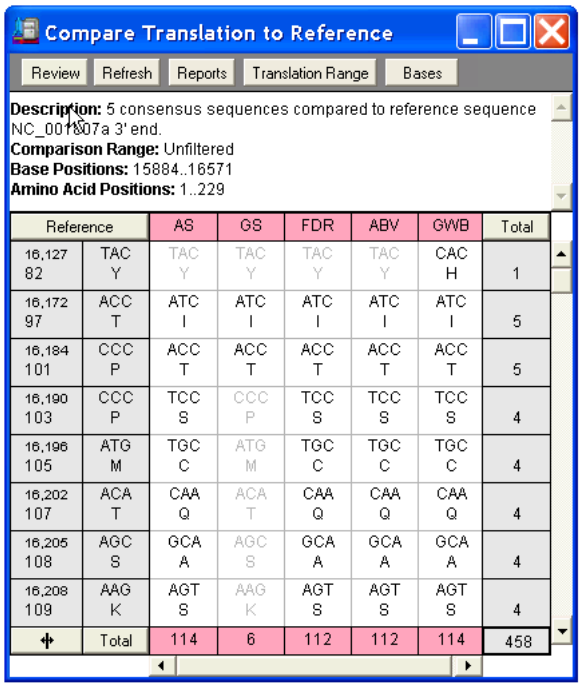
The Translated consensus Variance Table

Sequencher allows you to compare the translation of a consensus sequence to the translation of a Reference Sequence. Go to the **Contig** menu and click on **Compare Translation to Reference.**

You can also use the **Compare Translation to Reference** command to compare the translations of consensus sequences from multiple contigs to the translation of their common Reference Sequence.. These contigs are formed when you use the **Assemble By Name** with a Reference Sequence command.

For both of these operations, select one or more contigs in the Project Window that share a common Reference Sequence. Then, from the **Contig** menu, choose **Compare Translation to Reference.** The Translated consensus Variance Table appears in a new window (see Figure 13-7).

Figure 13-7 Translated Variance Table for consensus sequences from different contigs



Reference		AS	GS	FDR	ABV	GWB	Total
16,127 82	TAC Y	TAC Y	TAC Y	TAC Y	TAC Y	CAC H	1
16,172 97	ACC T	ATC I	ATC I	ATC I	ATC I	ATC I	5
16,184 101	CCC P	ACC T	ACC T	ACC T	ACC T	ACC T	5
16,190 103	CCC P	TCC S	CCC P	TCC S	TCC S	TCC S	4
16,196 105	ATG M	TGC C	ATG M	TGC C	TGC C	TGC C	4
16,202 107	ACA T	CAA Q	ACA T	CAA Q	CAA Q	CAA Q	4
16,205 108	AGC S	GCA A	AGC S	GCA A	GCA A	GCA A	4
16,208 109	AAG K	AGT S	AAG K	AGT S	AGT S	AGT S	4
	Total	114	6	112	112	114	458

Setting the Variance Table Conditions

Restricting the Comparison Range

You can restrict the Comparison Range by Feature Key if you have any Features in your exemplar. Construct a Variance Table then go to the button bar and click on the **Comparison Range** button. The **Comparison Range** dialog window appears (see Figure 13-8).

In the default setting the **Unfiltered** radio button is checked. Leave this radio button checked if you decide to leave your table unfiltered.

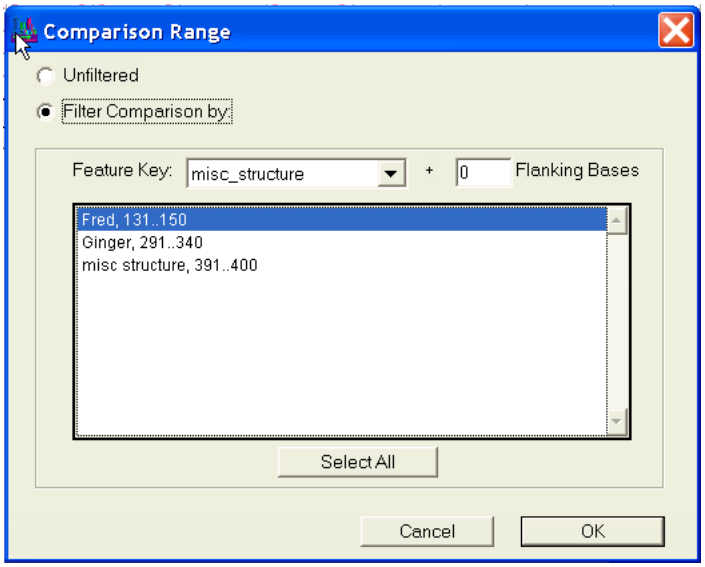
If you want to restrict the Comparison Range, click the **Filter by** radio button. Once you have clicked the **Filter by** radio button the **Feature Key:** drop down menu becomes active. You can now choose which Filter Key to use by making a selection from this menu.

Once you have selected a Feature Key, you will see a list of features annotated with that key in the pane located below the **Feature Key:** drop down menu. You can select a single feature, a continuous range or a discontinuous range of features. You can also select all the features within this pane by clicking on the **Select All** button.

You can widen your Comparison Range by typing a number into the **Flanking Bases** box. For example, by typing the number ten in this box, the Comparison Range will include 10 flanking bases on either side of the chosen Feature Key.

Once you have made your selection click the **OK** button. The Variance Table will be rebuilt automatically.

Figure 13-8 Comparison Range dialog window



Restricting the Translation Range

Reference Sequences are often marked up with GenBank-style features. (See Chapter 17 “Motifs and Features” for further information.

You can reduce the scope of your Translated Variance Table to a single feature. Some features such as mRNA or CDS may be composed of several elements. These elements are concatenated and treated as a single feature. Translation of the DNA sequence starts in the first frame, with the first base in the feature range. Figure 13-9 illustrates an unfiltered Comparison Range, a joined feature (blue) and a single feature (pink).

Figure 13-9 Restricting the Comparison Range using Features

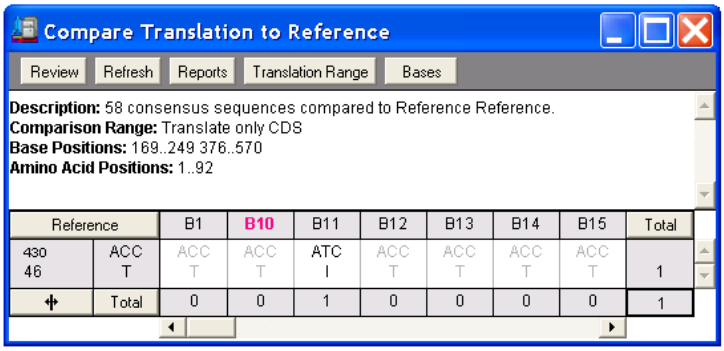
View your Translated Variance Table and go to the button bar and click on the **Translation Range** button. The **Translation Range** dialog window appears. You will see that the **Translate Entire Sequence** radio button is checked. This is your default. Leave this radio button checked if you decide to leave your table unfiltered.

If you want to restrict the Comparison Range, click the **Filter Translation by:** radio button. Once you have clicked the **Filter Translation by:** radio button the **Feature Key:** drop down menu becomes active. You can now choose which Filter Key to use by making a selection from this menu.

Once you have selected a Feature Key, you will see a list of features annotated with that key in the pane located below the **Feature Key:** drop down menu. You can select a single feature, a continuous range or a discontinuous range of features. You can also select all the features within this pane by clicking on the **Select All** button.

Once you have made your selection click the **OK** button. The Translated Variance Table will be rebuilt automatically (see Figure 13-10).

Figure 13-10 Translated Variance Table restricted by Translation Range

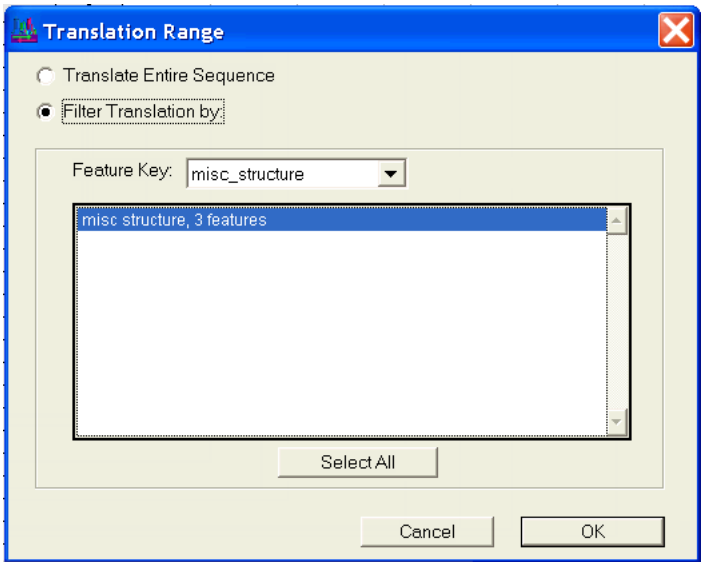


The table header indicates how the Comparison Range was restricted. For some Feature Keys the actual feature is composed of several components joined together (see Figure 13-9). The Base Positions indicate the exact range of bases used for each component.

Features which share the same Feature Key, such as CDS, and which are joined (the Feature has a Join qualifier) will appear in the Feature Listing as CDS [01], CDS [02]. The same Feature will appear in the Translation Range dialog window as a single item (see Figure 13-11).

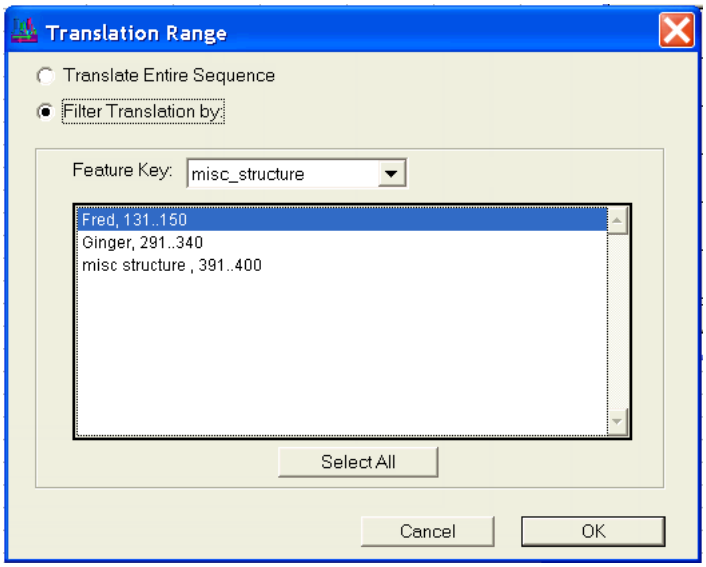
See Chapter 17 “Motifs and Features” and Appendix 28 “Feature Keys and Qualifiers” for more information.

Figure 13-11 Feature key with more then one component



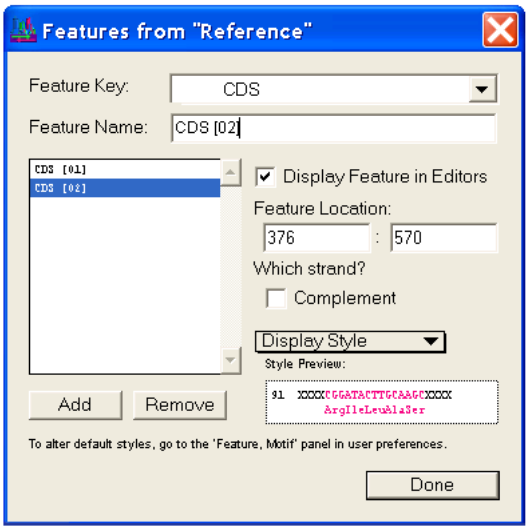
If you prefer to have each component of a Joined Feature as a separate item you must give them different names. You can do this by editing the Feature using the **Edit Features…** command (see figure below).

Figure 13-12 Appearance of features once they have been renamed



You can also create your own joined features. Use the **Edit Features…** command to create a series of features which share the same name followed by a bracketed number, CDS [01], CDS [02] (see Figure 13-13).

Figure 13-13 Creating a Joined feature



Setting the Variance Table Display Options

Throughout Sequencher, you can customize the display of your sequences using the commands from the **View** menu. When creating a Variance Table, you can also specify a number of display options (see Table 13-1).

Table 13-1 View Options

	Variance Table	consensus Variance Table	Translated Variance Table	Translated consensus Variance Table
Base Ambiguities	No	No	No	No
Base Edits	Yes	No	No	No
Display Color Bases	Yes	Yes	No	No
Display Features	Yes	Yes	No	No
Display Motifs	No	No	No	No
Colors As Backgrounds	Yes	Yes	No	No
Display Base Confidences	Yes	No	No	No
Labels	Yes	Yes	Yes	Yes

To highlight variants in your data, go to the **View** menu and click on **Display Base Confidences.** This will add a background color to the cells in your Variance Table to

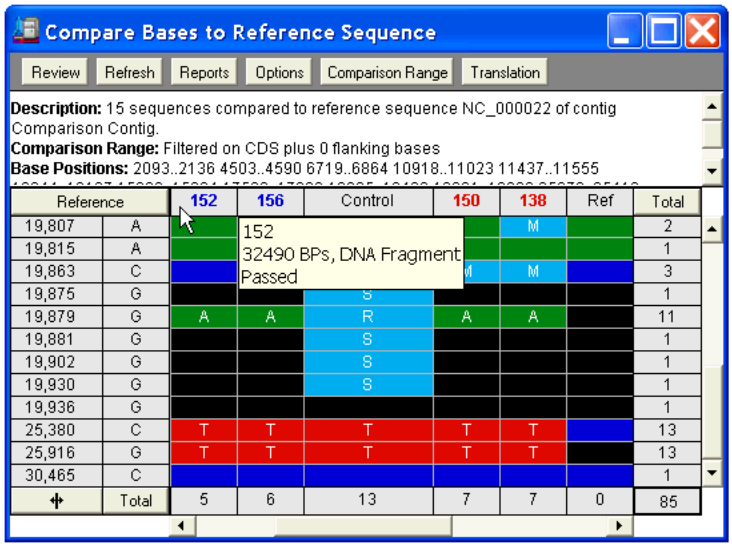
indicate that there is an associated confidence or quality score range for the variant. Variants that require additional checking because they have a low confidence or quality score will be highlighted with dark blue. Those with medium or high confidence scores will be highlighted with a medium or pale blue background.

If you click on **Display Color Bases** and **Display Colors as Backgrounds** it can help you to see patterns in your data (see Figure 13-14)**.** You can use **Base Ambiguities As** to distinguish ambiguous or edited bases. You can also toggle the display of features from the **View** menu.

You can use the **Label** command to mark sequences in the table. Set your preferences using the **Label & Name** preferences pane. Use the boxes grouped under **Available Labels** in the preference pane to choose label colors and descriptions for marking individual sequences. Select the sequence you want to label by clicking its column header. Go to the **Edit** menu and choose the **Label** submenu. The display name of the sequence will change to match the color of the **Label.** (See Figure 13-14.) Notice how the tool tip provides the name of the sample, its label and other attributes.

(Refer to Chapter 21 "Customizing Sequencher and User Preferences" for more detailed information.)

Figure 13-14 The Variance Table Display Options



Setting the Consensus Calculation for a Variance Table

You can generate a Variance Table where the exemplar is a consensus sequence rather then the Reference Sequence by using the **Compare Bases To** command from the **Sequence** menu. You generate a consensus Variance Table using the **Compare Consensus to Reference** command from the **Contig** menu.

In either case, the method you choose to calculate your consensus can affect your results. Generally, using the command **Consensus Inclusively** will report more variants than **Consensus by Plurality**.

Sequencher recalculates the consensus on the fly, so if you change the methods you are using to compute your consensus, you can reconstruct the Variance Table quickly. Just click on the **Refresh** button in the button bar at the top of the Variance Table to display the results of the new consensus calculation.

For more information on consensus calculations, see Chapter 11 "The Contig Editor".

Variance Table Options

If you need to change User Preferences for your Variance Table click the **Options** button in the Variance Table button bar. The User Preferences dialog opens in a new window. (See Chapter 21 "Customizing Sequencher and User Preferences" for more detailed information.)

Working with the Variance Table

Navigating a Variance Table

To navigate through any form of Variance Table, use the arrow keys to move across the rows or down the columns.

When you are working with any Variance Table and only want to focus on differences, use an arrow key with the **Alt key**. This will skip any cells that match the comparison sequence. If you try to move the cursor beyond the edge of the table, you will be alerted by a beep.

Note: If you are using the **Compare Consensus to Reference** command, you will see the consensus base(s) for each contig you have selected in the same position. If you are using the **Compare Translation to Reference** command you will be viewing the consensus codon for each contig you have selected in the same position.

Sorting the Variance Table

The Variance Table's **Sort** functions are useful when, for instance, you need to segregate clone sequences that perfectly match your Reference Sequence or cluster sequences of like alleles (see Figure 13-14). The Variance Table simultaneously displays differences and calculates summary information about your selections. The Table lists the total number of differences *for each sequence* at the bottom of that sequence's column. The Variance Table also displays the total number of differences *for each position* to the right of each row. There are two **Total** buttons, at the bottom left and top right corners of the Variance Table. These act as **Sort** buttons**.**

The **Total** button, on the bottom left of the table, sorts the sample columns so that the samples with the greatest number of differences move to the left. Click **Total** again, and the columns sort in the opposite order.

There are two ways to define the sort order for the rows. You can sort by total number of differences or you can sort by 5'-3'order. In the default, unsorted display of the Variance Table, rows are sorted by base position in ascending (5'- 3') order. When you click on **Total** in the top right corner of the Variance Table, the rows are rearranged so that those with the greatest number of differences sort to the top. Click the **Total** button again to reverse the order.

To restore the order by base position, click on the button in the top left corner of the Variance Table. Click here again to have the rows sort by position in descending (3'-5') order.

Note: The text on the sort button is context sensitive and will change depending on the **Compare Bases To** command used to generate the Variance Table.

Removing columns from the Variance Table

As you start to work with your Variance Table you may decide that some sequences are no longer of interest. The **Edit** menu's **Remove From Table** command allows you to delete sequences selectively from your Variance Table.

To select a continuous range of sequences, choose the first sequence by clicking on its column header and then, holding down the **Shift** key, select the last sequence in the same fashion. To select a discontinuous list of sequences, hold down the **Control** key while clicking on the sequences. Go to the **Edit** menu and choose the **Remove From Table** command.

Note: The **Remove From Table** command does not remove sequences from the underlying contig.

Review Mode

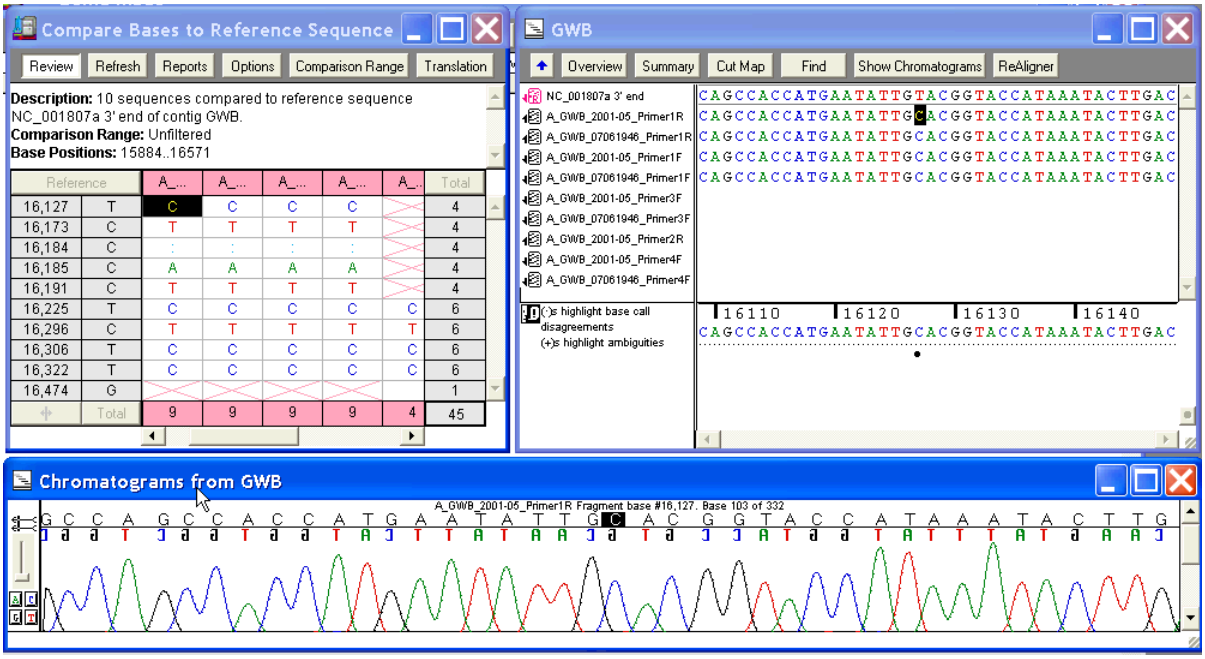
In **Review** mode, the Variance Table lets you view the data supporting each cell (see Figure 13-15). You can access the Review Mode by clicking on the **Review** button in the Variance Table button bar or by double clicking on any cell in the table. Sequencher will open both the Contig and Chromatogram editors with the associated base or column of bases selected.

You can specify the layout of these three Review windows, the Variance Table, the Contig Editor, and the Contig Chromatogram Editor, and then set the new organization as your default layout. After you have arranged your windows, go to the **Window** menu and choose the **Remember Window Layout** command, then select **Single Variance Table Review** from the submenu**.**

You can edit your data directly in the Variance Table in **Review** mode. Select the cell to be edited in the table, then type in your desired change. The change will be reflected in the Variance Table and the Contig Editor immediately.

Note: You can create features in the Variance Table by going to the **Sequence** menu and clicking on the **Mark Selection as Feature** command.

Figure 13-15 Comparing Sequences in Review Mode



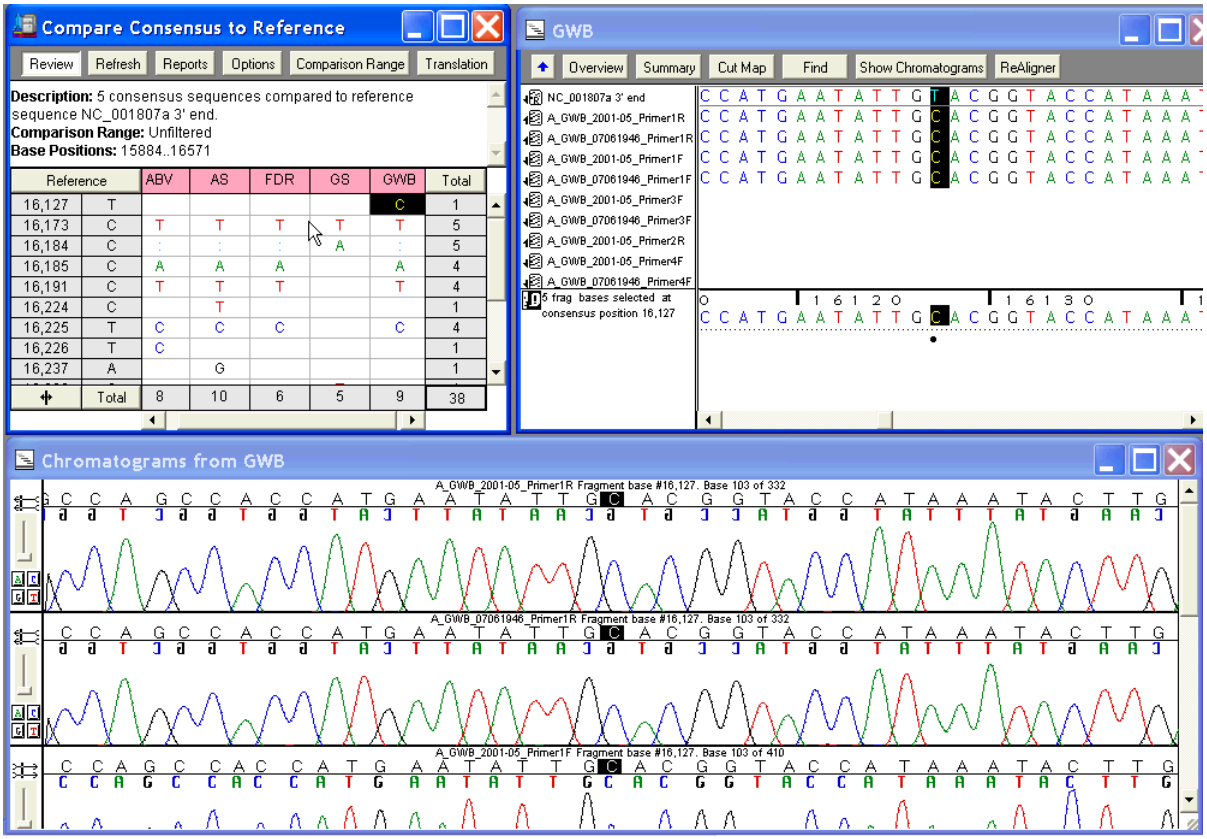
The Review Mode and Compare Consensus to Reference

The **Contig** menu's **Compare Consensus to Reference** command creates a Variance Table that reports differences from multiple contigs (see Figure 13-16). The **Review** mode lets you navigate efficiently between a difference in one contig and the equivalent position in the next. Each column in the Variance Table contains the differences from separate contigs. As you change your selection across the Variance Table rows, Sequencher closes the Contig Editor and the Contig Chromatogram Editor from your previous selection. It will then open the editors corresponding to your current selection.

Just as you can edit contigs from the consensus line in a contig editor, you can also edit contigs from a selection in the Variance Table **Review** mode. Your selection in the Variance Table corresponds to a selection in the consensus of the underlying contig.

Note: Remember that when you type over a base in the Variance Table, the underlying data in your contig will also change.

Figure 13-16 Comparing Contigs in Review Mode

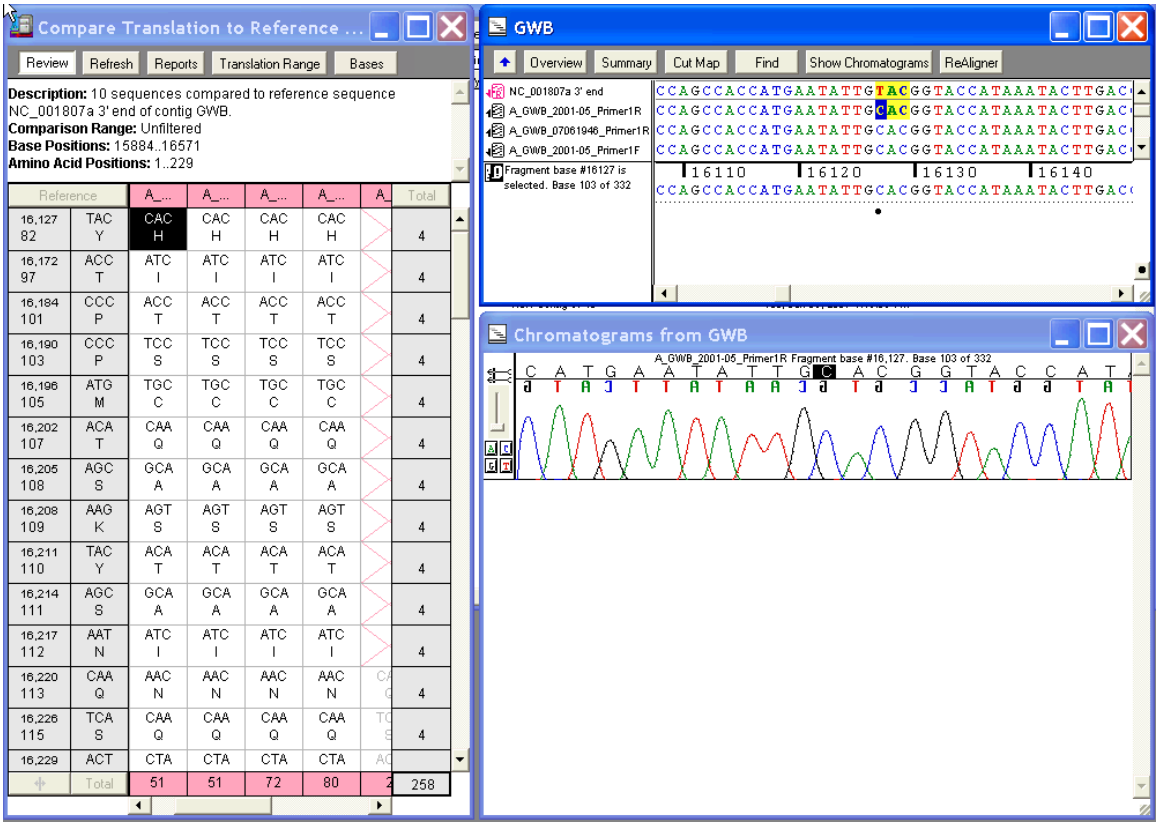


The Review Mode and the Translated Variance Table

The Translated Variance Table also has a **Review** button in its button bar. When you click this button you will see the associated Contig Editor and chromatograms (see Figure 13-17). When you click on an individual cell in the table, its codon is highlighted in the Contig Editor and the first base is selected. You will notice that the equivalent codon in the exemplar sequence is also highlighted.

Once you have arranged the individual windows to your satisfaction, you can save the layout by going to the **Window** menu and choosing **Remember Window Layout.** Then click on **Single Table Variance Review** from the submenu.

Figure 13-17 The Translated Variance Table in Review Mode



Note: You cannot edit your sequences directly from the Translated Variance Table (although you can do so in the ordinary Variance Table). Any edits to the Translated Variance Table must be made in the Contig Editor. In order to see the results of your changes in the Translated Variance Table, click on the **Refresh** button in the button bar.

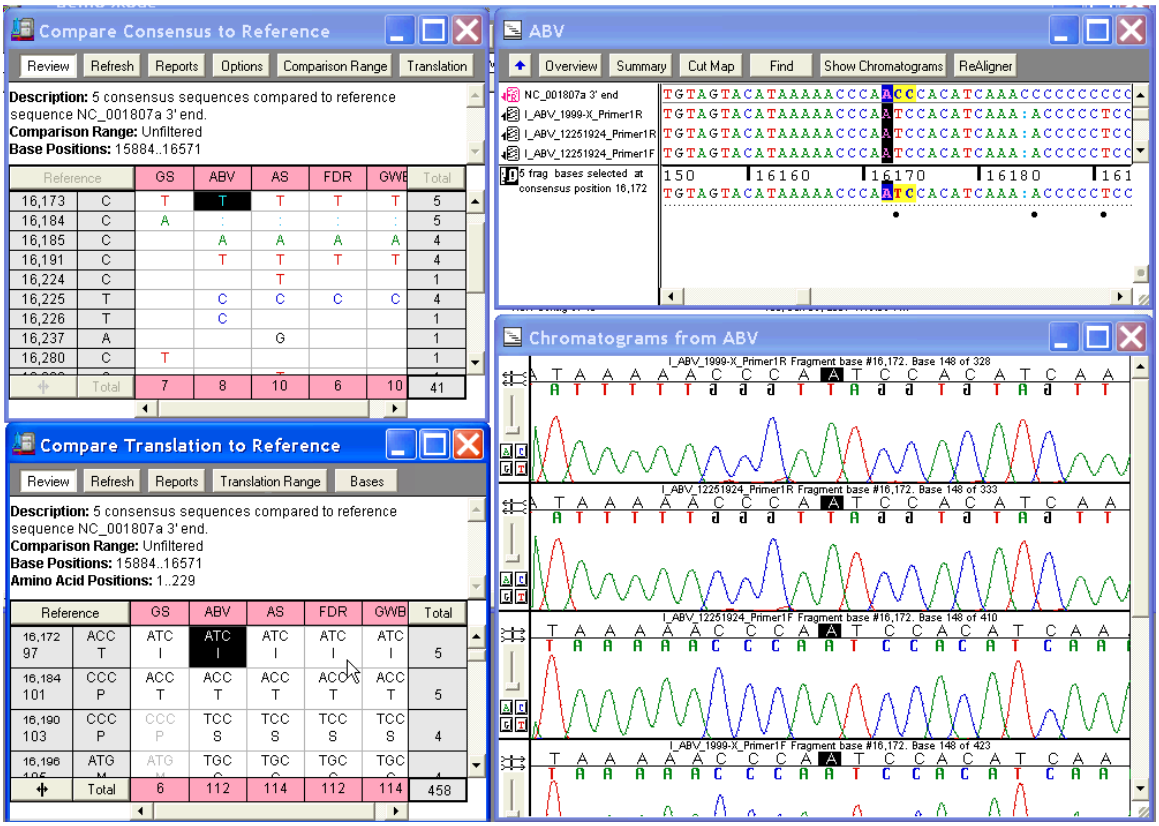
The Review Mode and Sister Tables

You can generate a "Sister Table" when you are working in an individual Variance Table. If you have generated a Variance Table you can see the Translated Variance Table for the same data set by going to the button bar and clicking on the **Translation** button (see Figure 13-18).

Similarly, if you have already generated a Translated Variance Table, you can see the Variance Table for the DNA sequences of the same data set by going to the button bar and clicking on the **Bases** button.

If you are in Review mode and have generated a Sister Table, you can arrange all four windows and save the layout by going to the **Window** menu and choosing **Remember Window Layou**t. Then click on **Double Table Variance Review** from the submenu.

Figure 13-18 Review mode showing Sister Tables



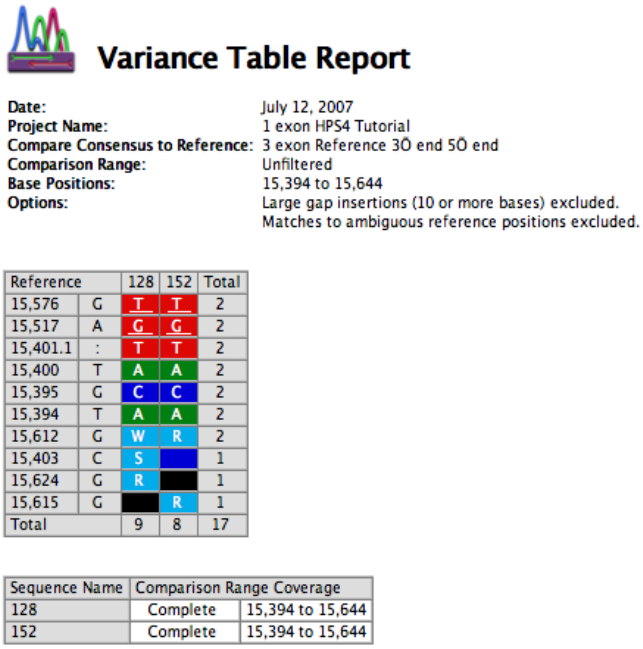
Variance Table Reports

The format of each Variance Table report is broadly similar. The report starts with a Common Header. This header includes the date, the project name, the type of table, Comparison Range, base positions and options. This is followed by a header containing more pertinent information and finally the body of the report.

Variance Table Report

The Variance Table Report consists of three parts: the Common Header, the Variance Table and the Comparison Range Coverage table. This table lists the name of each sequence, the type of coverage (complete or incomplete) and the range of bases in the comparison (see Figure 13-19).

Figure 13-19 Extract from Variance Table Report



Variance Table Report

Date: July 12, 2007
Project Name: 1 exon HPS4 Tutorial
Compare Consensus to Reference: 3 exon Reference 3Ō end 5Ō end
Comparison Range: Unfiltered
Base Positions: 15,394 to 15,644
Options: Large gap insertions (10 or more bases) excluded.
Matches to ambiguous reference positions excluded.

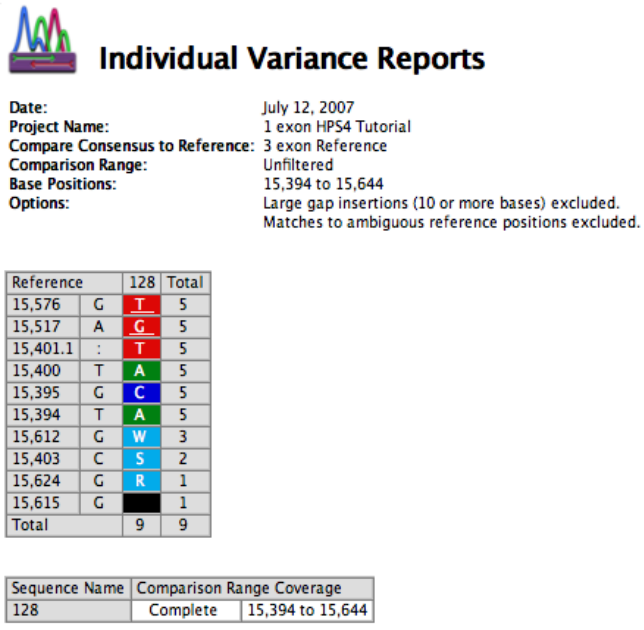
Reference		128	152	Total
15,576	G	T	T	2
15,517	A	G	G	2
15,401.1	:	T	T	2
15,400	T	A	A	2
15,395	G	C	C	2
15,394	T	A	A	2
15,612	G	W	R	2
15,403	C	S		1
15,624	G	R		1
15,615	G		R	1
Total		9	8	17

Sequence Name	Comparison Range Coverage	
128	Complete	15,394 to 15,644
152	Complete	15,394 to 15,644

Individual Variance Reports

You can create a separate report for each sample in the table. The Individual Variance Report lists the variants in a single column. This table is followed by the Comparison Range Coverage table which lists the name of each sequence, whether coverage is complete or incomplete, and the range of bases in the comparison (see Figure 13-20).

Figure 13-20 Extract from Individual Variance Report



Individual Variance Reports

Date: July 12, 2007
Project Name: 1 exon HPS4 Tutorial
Compare Consensus to Reference: 3 exon Reference
Comparison Range: Unfiltered
Base Positions: 15,394 to 15,644
Options: Large gap insertions (10 or more bases) excluded.
Matches to ambiguous reference positions excluded.

Reference		128	Total
15,576	G	T	5
15,517	A	G	5
15,401.1	:	T	5
15,400	T	A	5
15,395	G	C	5
15,394	T	A	5
15,612	G	W	3
15,403	C	S	2
15,624	G	R	1
15,615	G		1
Total		9	9

Sequence Name	Comparison Range Coverage	
128	Complete	15,394 to 15,644

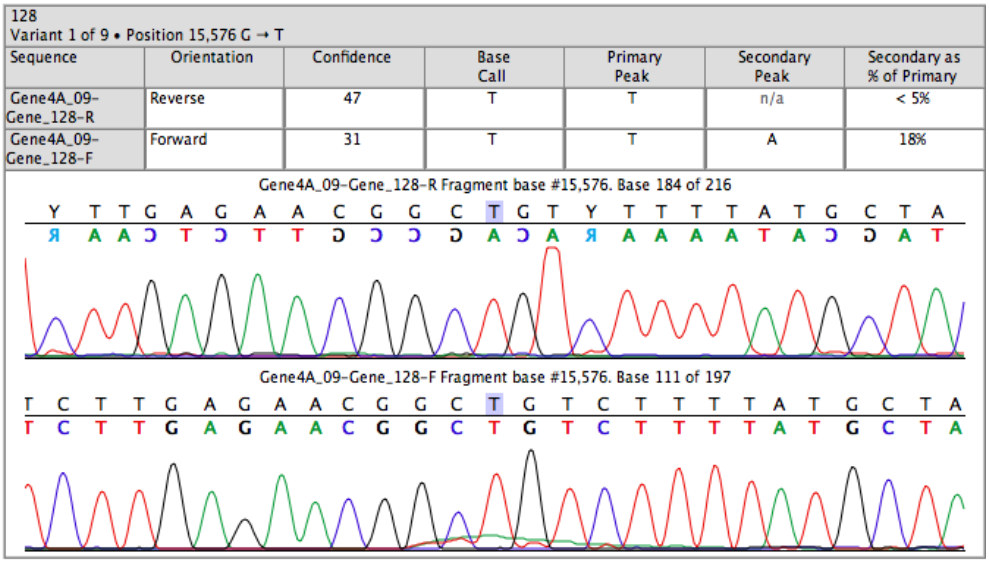
Note: This can produce a very long report.

Variance Detail Report

As its name indicates, the Variance Detail Report contains comprehensive information on the variants in a Variance Table. Below the Common Header, you will see detailed information for each of the described variants.

The information for each variant consists of the sample name and comparison range for the parent sequence. The Detail Table is divided into two parts. The first part contains the sequence name, orientation, confidence score (where available), base call, primary peak base call, secondary peak base call and the height of the secondary peak as a percentage of the primary peak. The second part of the table contains a tracelet, which is an extract of the sequence trace of the variant combined with some flanking bases (see Figure 13-21). The tracelet only appears in the report if the sequence has associated chromatogram data.

Figure 13-21 Extract from a Variance Detail Report



Note: If the Detail Table displays a gap for a specific base position, this will still be given a confidence value. The confidence value will be equal to the average confidence ratings of the bases flanking the gap. Sequencher will append an asterisk to the confidence value and a footnote will explain how the value was derived.

Population Report

The Population Report is a special report which groups together all sequences sharing the same set of variants. This report contains the Common Header and a report-specific header. The report-specific header lists the number of population groups and the how many samples this consists of. This header also indicates whether any samples have been dropped from the report. This will only happen if a sample does not have complete

Comparison Range coverage. This report consists of two types of tables, a Population Table and individual Detail Tables.

The Population Table lists the variants in each population group. It displays information on the number of samples which contain each variant base as well as the total number of samples in each population group. Each population group is given a name based alphabetically on the first member of the group (see Figure 13-22).

The Detail Table includes information on the frequency of the population group and, the location of the variants with reference to the exemplar. It also lists the names of the contributing samples.

Figure 13-22 Extract from a Population Report

Participating Data: 5 population groups consisting of 5 samples
Nonparticipating Data: 0 samples dropped from the report due to incomplete comparison range coverage

Populations

Reference		116	128	150	152	156	Total Samples
15,576	G	T	T	T	T	T	5
15,517	A	G	G	G	G	G	5
15,401.1	:	T	T	T	T	T	5
15,400	T	A	A	A	A	A	5
15,395	G	C	C	C	C	C	5
15,394	T	A	A	A	A	A	5
15,612	G		W		R	R	3
15,403	C	S	S				2
15,624	G		R				1
15,615	G				R		1
Total Samples		1	1	1	1	1	

116 – Like		
Frequency	20.00%	
Variants	7	15576 T, 15517 G, 15401.1 T, 15400 A, 15395 C, 15394 A, 15403 S
Samples	1	116

Working with Variance Table Reports

You now have a number of ways to present the data from your Variance Table. You can use the table's **Reports** function to print your data or to export it. The options vary depending on the data you have chosen to use in your report.

Printing a Variance Table Report

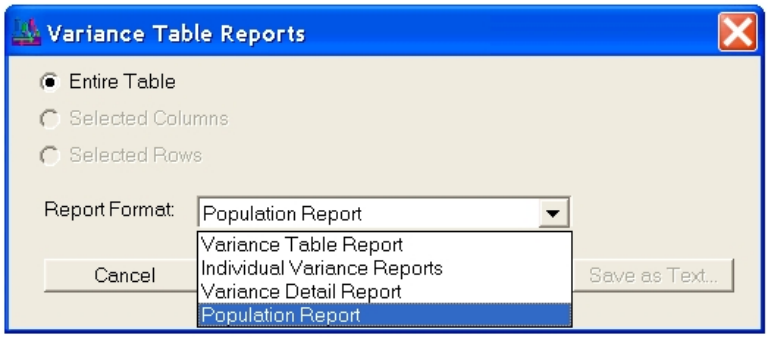
You can print the data from an entire table, selected columns or selected rows. The type of report you can generate will depend on this initial selection (see Table 13-2).

Table 13-2 Table 13 2 Variance Table print options

	Variance Table Report	Individual Variance Table Reports	Variance Detail Report	Population Report
Entire Table	Yes	Yes	Yes	Yes
Selected Columns	Yes	Yes	Yes	Yes
Selected Rows	Yes	n/a	Yes	n/a

To print all the data in a table, click on the **Reports** button. The **Variance Table Reports** dialog box appears (see Figure 13-23). A radio button called **Entire Table** will be checked. Below that you will see a drop down menu called **Report Format:**. Choose either **Variance Table Report**, **Individual Variance Table Reports, Variance Detail Report** or **Population Report**. Click the **Open Report…** button. The **Sequencher Report Viewer** opens in a new window.

Figure 13-23 Variance Table Reports drop down menu



Use the **Print Preview** button from the button bar at the top of the Report Viewer to preview your report (see Figure 13-24).

Figure 13-24 Reports Viewer button bar

The **Report View** is replaced by the **Preview View**. Note the **Zoom In** and **Zoom Out** buttons in the button bar (see Figure 13-25). Click the **Report View** button to return to the previous view.

Figure 13-25 Print Preview button bar

Use the **Page Setup** button to choose paper size. Click on the **Print** button to send the report to a printer. You can also save the report in PDF format by clicking the **Save as PDF…** button. If you wish to dismiss the window without performing any of these actions click on the **Close** button.

Printing selected data from a Variance Table

You can print selected data from your table. You have several options. You can remove sample sequences from the table using the **Remove from Table** command. You can print selected variants by choosing selected rows and you can print selected samples by choosing selected columns. You may already have restricted the data in the table by using the **Comparison Range** command.

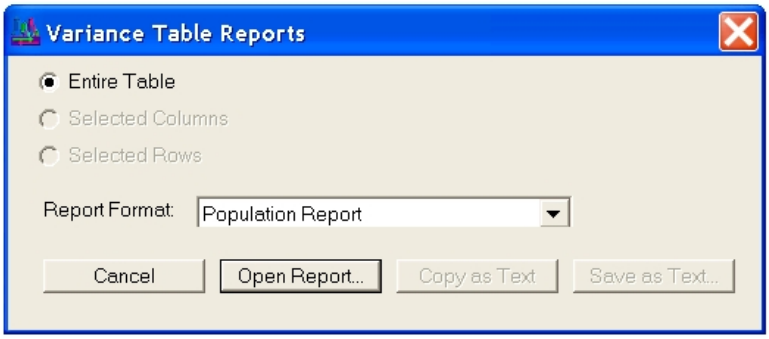
If you want to remove sample sequences from your table, select a continuous range of sequences, and choose the first sequence by clicking on its column header. Then, holding down the **Shift** key, select the last sequence in the same fashion. To select a discontinuous list of sequences, hold down the **Control** key while clicking on the sequences. Go to the **Edit** menu and choose the **Remove From Table** command.

Note: This does *not* remove sequences from the underlying contig.

If you want to print selected samples from your table you can choose a continuous or a discontinuous range of sequences by clicking in the column header as described previously.

Click on the **Reports** button. The **Variance Table Reports** dialog box appears (see Figure 13-26). A radio button called **Selected Columns** will be checked. Below that you will see a drop down menu called **Report Format:**. Choose either **Variance Table Report**, **Individual Variance Table Reports, Variance Detail Report** or **Population Report**. Click the **Open Report…** button. The **Sequencher Report Viewer** opens in a new window.

Figure 13-26 Variance Table Reports dialog window



Use the **Print Preview** button from the button bar at the top of the Report Viewer to preview your report. Click on the **Print** button to send the report to a printer.

If you want to print selected variants from your table you can choose a continuous or a discontinuous range of base positions by clicking the equivalent exemplar base position as described previously.

Click on the **Reports** button. The **Variance Table Reports** dialog box appears (see Figure 13-26). A radio button called **Selected Rows** will be checked. Below that you will see a drop down menu called **Report Format:**. Choose either **Variance Table Report** or **Variance Detail Report**. Click the **Open Report…** button. The **Sequencher Report Viewer** opens in a new window.

Use the **Print Preview** button from the button bar at the top of the Report Viewer to preview your report. Click on the **Print** button to send the report to a printer.

Note: Not all reports are available when you print selected rows.

Exporting a Variance Table Report

You can export the data from an entire table, selected columns or selected rows. The type of report you can generate will depend on this selection (see Table 13-3).

Table 13-3 Variance Table Report export options

	Variance Table Report	**Individual Variance Table Reports**	**Variance Detail Report**	**Population Report**
Entire Table	Yes	Yes	No	No
Selected Columns	Yes	Yes	No	No
Selected Rows	Yes	n/a	No	n/a

You can remove sample sequences from your table before you export your data. Select a continuous range of sequences, choose the first sequence by clicking on its column header. Then, holding down the **Shift** key, select the last sequence in the same fashion. To select a discontinuous list of sequences, hold down the **Control** key while clicking on the sequences. Go to the **Edit** menu and choose the **Remove From Table** command.

Note: This does *not* remove sequences from the underlying contig.

You can then export the remaining data as normal.

To export all the data in the table, click on the **Reports** button. The **Variance Table Reports** dialog box appears. A radio button called **Entire Table** will be checked. Below that you will see a drop down menu called **Report Format:**. Choose either **Variance Table Report** or **Individual Variance Table Reports**. If you choose **Variance Table** Sequencher will generate a single table. If you choose **Individual Variance Table Reports** each sample will appear in a separate report.

There are four buttons below the **Report Format** drop down menu, **Cancel, Reports, Copy as Text** and **Save as Text... Save as Text…** is the default setting**.** If you click on the **Save as Text…** button, Sequencher presents you with a dialog to assign the name and location of your export. The report is saved in a tabbed text format.

Note: You can select and then remove columns from your table using the **Remove From Table** command before you export your data.

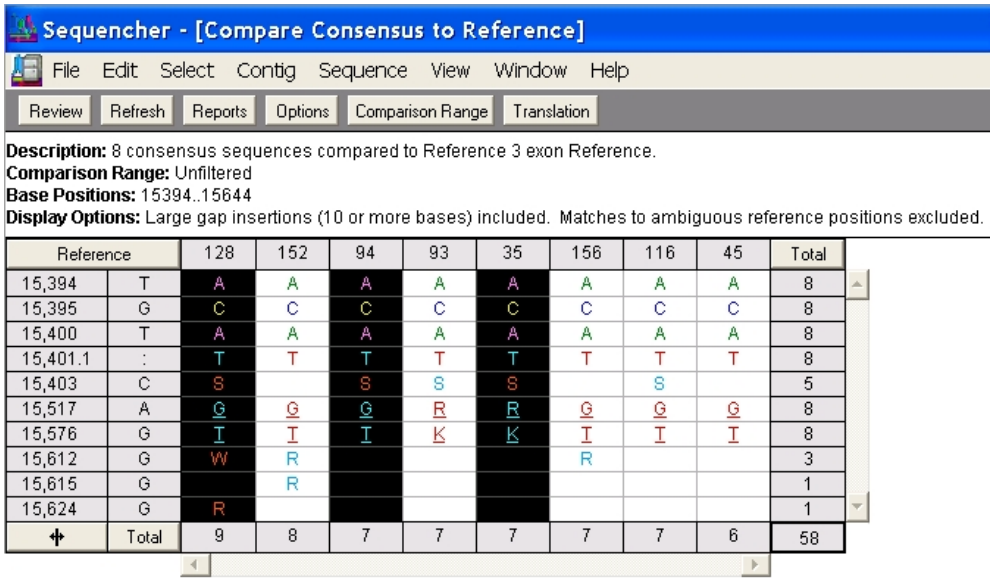
Exporting selected sample sequences

Each column in the table represents a sequence or consensus. If you need to export specific columns in the table, make your selection by using **Shift click** for a continuous range or **Ctrl click** for a discontinuous range (see Figure 13-27). Then click on the **Reports** button. The **Variance Table Reports** dialog box appears, the **Selected Columns** radio button is checked. Choose a **Report Format** from the drop down menu.

Select **Individual Variance Reports** if you want to save each column (sample) as a separate file. Select **Variance Table Report** if you want the selected columns in one table. Click the **Save as Text…** button, Sequencher presents you with a dialog to assign the name and location of your export.

You can copy your Variance Table into another document or application that accepts tabbed text. To paste the contents of your Variance Table Report in the document of your choice, click on the **Copy as Text** button. The data can now be pasted into its new location.

Figure 13-27 Variance Table with discontinuous range of columns selected


Sequencher - [Compare Consensus to Reference]
File Edit Select Contig Sequence View Window Help
Review Refresh Reports Options Comparison Range Translation
Description: 8 consensus sequences compared to Reference 3 exon Reference.
Comparison Range: Unfiltered
Base Positions: 15394..15644
Display Options: Large gap insertions (10 or more bases) included. Matches to ambiguous reference positions excluded.

Reference		128	152	94	93	35	156	116	45	Total
15,394	T	A	A	A	A	A	A	A	A	8
15,395	G	C	C	C	C	C	C	C	C	8
15,400	T	A	A	A	A	A	A	A	A	8
15,401.1	:	T	T	T	T	T	T	T	T	8
15,403	C	S		S	S	S		S		5
15,517	A	G	G	G	R	R	G	G	G	8
15,576	G	T	T	T	K	K	T	T	T	8
15,612	G	W	R				R			3
15,615	G		R							1
15,624	G	R								1
	Total	9	8	7	7	7	7	7	6	58

Exporting data for selected variants

Each row in the table represents a variant. If you need to export selected rows, make your selection by using **Shift click** for a continuous range or **Ctrl click** for a discontinuous range. Then click on the **Reports** button. The **Variance Table Reports** dialog box appears. The **Selected Rows** radio button is checked. There are four buttons, **Cancel, Open Report…, Copy as Text** and **Save**. **Save as Text…** is the default setting.

The chosen rows are saved as one table. The options in the **Reports Format:** drop down menu are limited to **Variance Table Report** and **Variance Detail Report.** Click the **Save as Text…** button. You will see a dialog window prompting you to assign the name and location of your export. You can type a name for the table in this window. Click the **Save** button to save your table as a tabbed text file.

Although the **Selected Rows** radio button is checked, you can change your selection to include all rows by clicking the **Entire Table** radio button.

You can copy your Variance Table in to another document or application that accepts tabbed text. To paste the contents of your Variance Table Report in the document of your

choice, click on the **Copy as Text** button. The data can now be pasted into its new location.

Translated Variance Table Report

Translated Variance Table Report

The Translated Variance Table Report consists of three parts: the header information, the Comparison Range Coverage and the Translated Variance Table.

The Translated Variance Table contains the name of the sample sequence, the variant codon, the amino acid residue and the type of coverage (whether the coverage is complete or incomplete) for each variant in the table. The Comparison Range Coverage table lists the name of each sequence and the range of bases in the comparison.

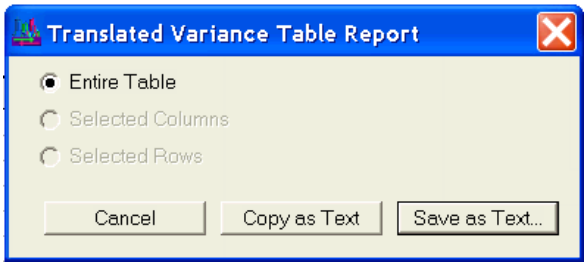
Working with a Translated Variance Table Report

There are a number of ways to present the data from a Translated Variance Table. You can export the data from an entire table, selected columns or selected rows. You can also remove data from the table before you export it.

Exporting a Translated Variance Table Report

To export all the data in the table, click on the **Reports** button. The **Translated Variance Table Reports** dialog box appears (see Figure 13-28). A radio button called **Entire Table** will be checked.

Figure 13-28 Translated Variance Table Reports dialog window



You will see three enabled buttons in the **Translate Variance Table Reports** dialog window, **Cancel, Copy as Text** and **Save as Text...**

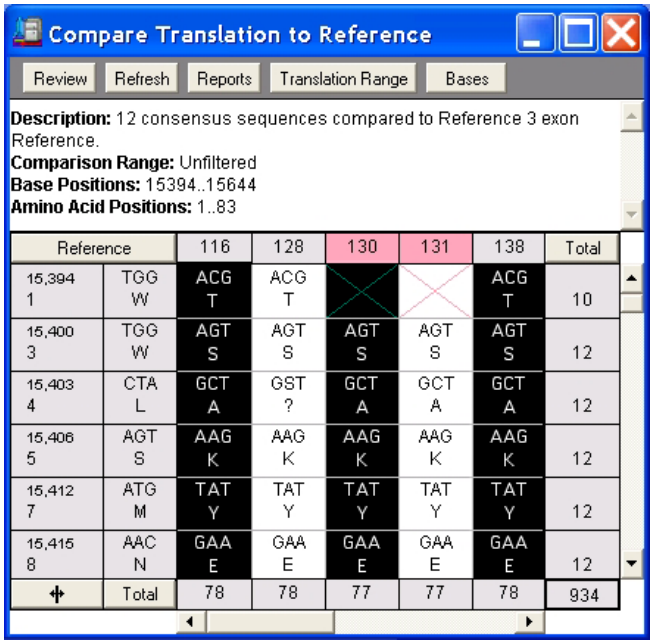
Save as Text… is the default setting. If you click on the **Save as Text…** button, Sequencher presents you with a dialog to assign the name and location of your export. The report is saved in a tabbed text format.

If you prefer to copy your Translated Variance Table to another document or application which accepts tabbed text, click on the **Copy as Text** button. The data can now be pasted into its new location.

Exporting selected sample sequences

Each column in the table represents a sequence or consensus. If you need to export specific columns in the table, make your selection by using **Shift click** for a continuous range or **Ctrl click** for a discontinuous range (see Figure 13-29). Then click on the **Reports** button. The **Translated Variance Table Reports** dialog box appears (see figure above). A radio button called **Selected Columns** will be checked.

Figure 13-29 Discontinuous selection of columns


Compare Translation to Reference
Review
Refresh
Reports
Translation Range
Bases
Description: 12 consensus sequences compared to Reference 3 exon Reference.
Comparison Range: Unfiltered
Base Positions: 15394..15644
Amino Acid Positions: 1..83
Reference
116
128
130
131
138
Total
15,394 1 TGG W ACG T ACG T ACG T 10
15,400 3 TGG W AGT S AGT S AGT S AGT S AGT S 12
15,403 4 CTA L GCT A GST ? GCT A GCT A GCT A 12
15,406 5 AGT S AAG K AAG K AAG K AAG K AAG K 12
15,412 7 ATG M TAT Y TAT Y TAT Y TAT Y TAT Y 12
15,415 8 AAC N GAA E GAA E GAA E GAA E GAA E 12
Total 78 78 77 77 78 934

Choose one of the three enabled buttons in the **Translated Variance Table Reports** dialog box, **Cancel, Copy as Text** and **Save as Text... Save as Text…** is the default setting**.** If you click on the **Save as Text…** button, Sequencher presents you with a dialog to assign the name and location of your export. The report is saved in a tabbed text format.

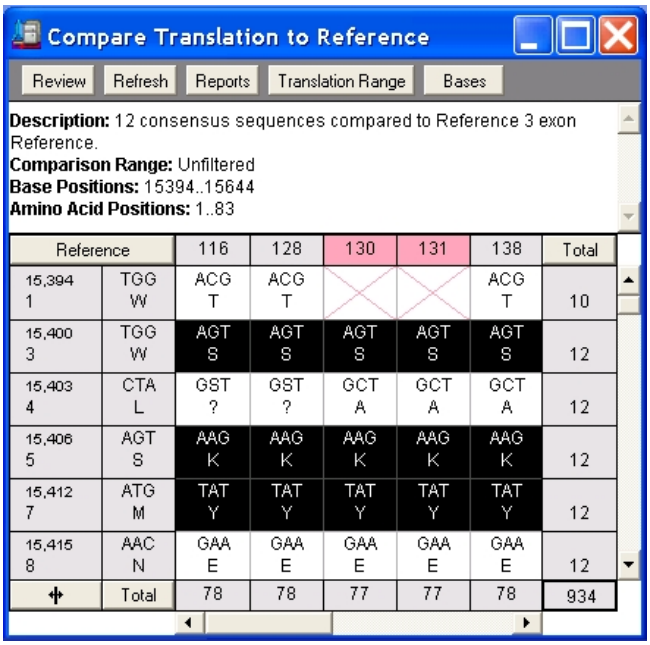
Although the **Selected Columns** radio button is checked, you can change your selection to include all rows by clicking the **Entire Table** radio button.

You can copy your Translated Variance Table into another document or application that accepts tabbed text. To copy the contents of your Translated Variance Table Report, click on the **Copy as Text** button. The data can now be pasted into its new location.

Exporting data for selected variants

Each row in the table represents a variant. If you need to export selected rows, make your selection by using **Shift click** for a continuous range or **Ctrl click** for a discontinuous range (see Figure 13-30). Then click on the **Reports** button. The **Translated Variance Table Reports** dialog box appears. The **Selected Rows** radio button is checked. Three buttons are enabled, **Cancel, Copy as Text** and **Save as Text. Save as Text…** is the default setting**.**

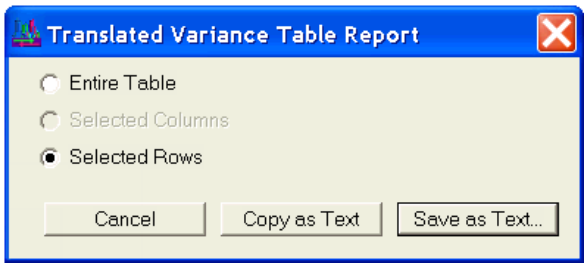
Figure 13-30 Translated Variance Table - selected variants


Compare Translation to Reference
Review
Refresh
Reports
Translation Range
Bases
Description: 12 consensus sequences compared to Reference 3 exon Reference.
Comparison Range: Unfiltered
Base Positions: 15394..15644
Amino Acid Positions: 1..83
Reference
116
128
130
131
138
Total
15,394 1 TGG W ACG T ACG T ACG T 10
15,400 3 TGG W AGT S AGT S AGT S AGT S AGT S 12
15,403 4 CTA L GST ? GST ? GCT A GCT A GCT A 12
15,406 5 AGT S AAG K AAG K AAG K AAG K AAG K 12
15,412 7 ATG M TAT Y TAT Y TAT Y TAT Y TAT Y 12
15,415 8 AAC N GAA E GAA E GAA E GAA E GAA E 12
Total 78 78 77 77 78 934

The rows you have selected will be saved in one table. Click the **Save as Text…** button. A dialog window to assign the name and location of your export appears. You can type a name for the table in this window. Click the **Save** button to save your table as a tabbed text file.

Although the **Selected Rows** radio button is checked, you can change your selection to include all rows by clicking the **Entire Table** radio button (see Figure 13-31).

Figure 13-31 Translated Variance Table Reports dialog window - Selected Rows



You can copy your Translated Variance Table into another document or application that accepts tabbed text. To paste the contents of your Translated Variance Table Report click on the **Copy as Text** button.

Removing data from a table before export

If you need to reduce the amount of data in your table you can remove selected sample sequences.

Make your selection by using **Shift click** for a continuous range or **Ctrl click** for a discontinuous range.

Go to the **Edit** menu and choose the **Remove From Table** command. Your selection will be removed. You can then export the remaining data as usual.

Note: This command does *not* remove sequences from the underlying contig.

14. The Summary Report

In this chapter we explain how to use the Summary Report. This is a static report which has a number of options making it a useful addition to the Variance Table for mutation hunting, verifying re-sequenced fragments, comparisons to a Reference Sequence or just as a report for your lab notebook.

The Summary Report view

View by Summary

You can display this report by going to the **Contig Editor** and clicking on the **Summary** button in either the **Bases View** window or the **Overview…**. Alternatively, you can go to the **View** menu, choose **Bases, Map, Overview…** and then click on **Summary Report.**

You will be presented with a new window which shows the assembled sequences within your chosen contig. There are five buttons arranged in a button bar (see below).

Button	Description
Bases	This button returns you to the contig editor
Overview	This returns you to the contig overview
Ruler	This button brings up a simple text editor-like ruler
Options	This controls which information will be displayed
Find	This button brings up a window into which you can type a substring and perform a search (see Chapter 18 “Finding Items”).

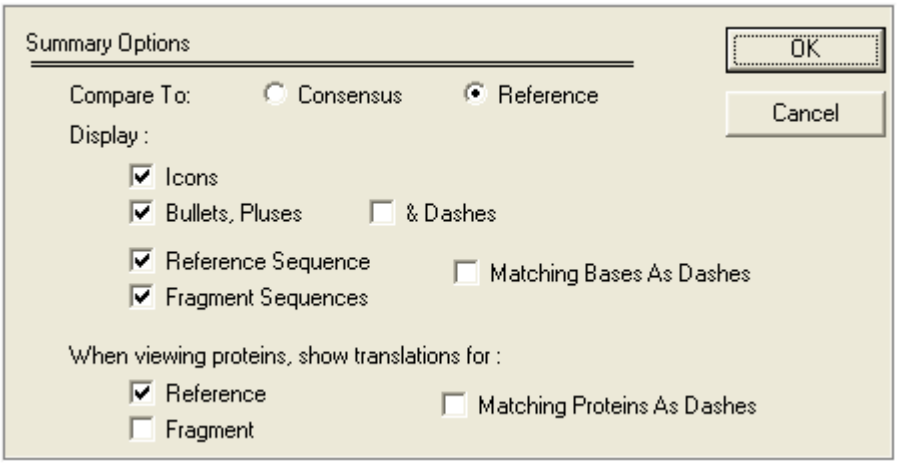
Summary report display options

Sequencher has a default format for the summary report but you can change it to display different types of information. Click the **Options** button or go to the **View** menu and choose **View Options…**

The **Summary Options** are divided into those that affect the sequence fragments and those that affect the display of protein translations. You can change the consensus calculation while in this view. Any change you make will be reflected immediately in the report.

Sequencher displays the window shown in Figure 14-1.

Figure 14-1 Summary report options



Compare To Consensus or Reference Sequence

If you have a Reference Sequence in your contig you can choose whether to display a **Summary Report** compared to the consensus sequence or to the assembled Reference Sequence. The default setting is **Compare To: Consensus** but you can change this by clicking on the **Reference** radio button. Some of the **Display** options are context sensitive and will change to reflect your choice.

The **Reference** radio button will be grayed out when the selected contig does not contain a Reference Sequence.

Icons

If you check the box for the **Icons** display, you will see the sequence fragment icons positioned to the left of the fragment name. Unchecking this option will hide the icons.

Note: When you have a Reference Sequence in your contig its icon is distinguished by the addition of an “R”.

SEQUENCHER® User Manual © 1991 – 2011 Gene Codes Corporation, Inc. All rights reserved.

Bullets, Pluses and Dashes

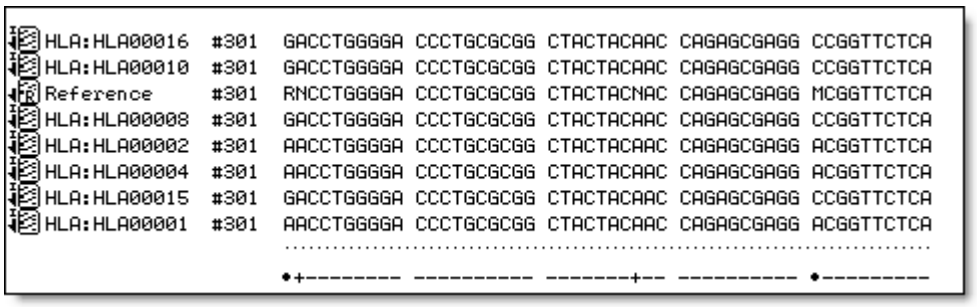
Clicking the **Bullets, Pluses** boxes highlights disagreements and ambiguities in a line below the consensus. If you also click on **& Dashes** you are marking positions where the bases in individual fragments match so that the disagreements are more visible.

Note: You can display **Bullets, Pluses** without the **& Dashes** but you cannot display **& Dashes** without the **Bullets, Pluses**. However, see **Matching Bases As Dashes** below.

Consensus Sequence

You can choose to hide the consensus sequence by clicking on the **Consensus Sequence** box. Figure 14-2 shows a summary line without the consensus line but with **Bullets, Pluses** and **& Dashes** all checked.

Figure 14-2 Summary report with no consensus line, with bullets, pluses and dashes


HLA:HLA00016 #301 GACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG CCGGTTCTCA
HLA:HLA00010 #301 GACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG CCGGTTCTCA
Reference #301 RNCCTGGGGA CCCTGCGCGG CTACTACNAC CAGAGCGAGG MCGGTTCTCA
HLA:HLA00008 #301 GACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG CCGGTTCTCA
HLA:HLA00002 #301 AACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG ACGGTTCTCA
HLA:HLA00004 #301 AACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG ACGGTTCTCA
HLA:HLA00015 #301 GACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG CCGGTTCTCA
HLA:HLA00001 #301 AACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG ACGGTTCTCA

Fragment Sequences

You can choose to hide the fragment sequence by clicking on the **Fragment Sequences** box. This gives you a very compressed report that can be useful when viewing protein translations. Figure 14-2 above shows a summary line without the consensus line, but with **Bullets, Pluses** and **& Dashes** all checked.

Matching Bases as Dashes

Matching Bases As Dashes is a very powerful option when you are interested in any kind of mutation detection. When you check this option all the bases that match each other are converted to dashes.

If you have selected **Display Color Bases** the dashes will be marked in the color of the original base call. If you also go to the **View** menu and chose **Colors as Background,** each dash will appear embedded in a small color block. Bases that disagree are displayed as characters and as such stand out in stark relief. With this option, you can easily see mutation hot spots or clusters and potential SNPs.

Protein translations and the Summary Report

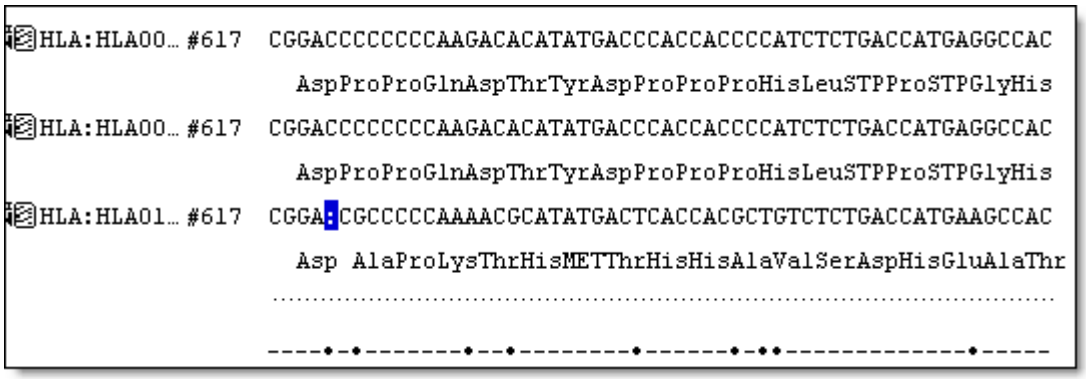
Sequencher can show protein translations for individual fragments and mark matching amino acids with dashes. This helps you see whether a particular discrepancy in an alignment has functional implications or whether it may simply be a silent mutation. If you combine options, you can turn off the DNA display altogether and just focus on your translations. Alternately, you can use a mixture of options to tailor the report to your specific needs.

Enabling the protein display

To enable the protein display in the summary view, go to the **View** menu and click on **Translation.** You can then choose from the following list of menu sub-options: **Single Stranded, Double Stranded, Protein First Frame, Protein Second Frame, Protein Third Frame** and**, Protein All Three Frames**.

To see the protein translations for the opposite strand go the **View** menu and select **Reverse & Comp**. Then select the desired option. Where the assembly algorithm has inserted a gap in a fragment Sequencher will translate the codon correctly.

Figure 14-3 Summary Report displaying translation across a sequence gap



Protein translations for the Consensus or Reference Sequence

This option is context sensitive. If you do not have a Reference Sequence or you have selected the **Consensus** radio button you will be able to show protein translations for the consensus. Click on the **Consensus** check box at the bottom of the **Summary Options** dialog window.

If you have a Reference Sequence in your contig and you have selected the **Reference** radio button you will be able to show translations for the Reference Sequence. Click on the **Reference** check box at the bottom of the **Summary Options** dialog window.

Protein translations for Fragments

To show protein translations for fragments click on the **Fragment** check box in the **Summary Options** dialog box.

Matching Proteins As Dashes

To show **Matching Proteins As Dashes** click on the check box in the **Summary Options** dialog box.

Using the formatting ruler

You can use the formatting ruler to change the blocking (for example, to groups of 3s instead of 10s,) or to show protein translations. (See Chapter 8 “The Sequence Editor” for more information on the formatting ruler.)

Click the **Ruler** button in the button bar or go to the **View** menu and choose **Display Format Ruler**. You can use the smaller font options in conjunction with the landscape page format to obtain full-width displays of your assembled sequences for use in posters and figures.

Reverse and Complement a contig

You can change the orientation of the contig while displaying any of its views by going to the **View** menu and choosing **Reverse & Comp.** You can display the contig’s original orientation by choosing **Reverse & Comp** again.

15. Chromatograms

In this chapter you will learn how to use the chromatograms from an automated sequencer to help you when editing your contigs. We discuss how Sequencher displays traces and secondary peaks to find heterozygotes and how you can edit bases from traces or revert to experimental data.

Working with automated sequencer data

When you import a trace file from an automated sequencer, Sequencher imports the chromatograms and the confidence values if these are available.

Note: If you import a text file containing just ASCII characters, you will not have imported traces. To import traces you must import the trace file. Text files from an automated sequencer will be only a couple of kilobytes in size, whereas trace files will be well over 50 kilobytes.

When you copy a Sequencher project file from one machine to another, you don't need to copy the trace files separately because Sequencher stores the trace information along with everything else in the project file. Sequencher compresses the imported traces to conserve disk space, so don't be alarmed if a project file containing several sequences from an automated sequencer is smaller than any one of the trace files.

Viewing chromatograms from a Sequence Editor

Sequencher's chromatogram display is tightly integrated with its other editors. You can view the entire original trace for a sequence fragment by opening the Sequence Editor for that sequence. Click the **Show Chromatogram** button in the button bar or go to the **Window** menu and choose **Chromatogram.** You can scroll either vertically (the default) or horizontally. To change orientation, click the appropriate button in the left bottom corner of the fragment chromatogram window.

Note: Remember that even if a sequence fragment is incorporated into a contig, you can still open its editor by double-clicking the sequence name in the list at the left of the contig editor.

Viewing chromatograms from a contig editor

The Sequencher contig editor allows you to view simultaneously all chromatogram data relevant to a consensus base call. If you are at the 5' end of the contig, Sequencher selects the most 5' base in the consensus and displays all the associated traces. If you are elsewhere in the contig, Sequencher selects the center-most consensus base in the contig display and displays the associated traces. This allows you to check the signal strength of any peak in conjunction with its base call and edit your data accordingly.

Select a contig column by clicking on its base in the consensus line (Figure 15-1). Then click the **Show Chromatogram** button or go to the **Window** menu and choose **Chromatogram.** Sequencher opens a window to display the traces (Figure 15-1). The bases you highlighted in the contig editor will also be highlighted in the trace windows.

Figure 15-1 Selected column chromatograms

When you click the **Show Chromatograms** button in a region without trace data you will see a "No Chromatogram Data" message in the Contig Chromatogram window. When you move the cursor back into a region where there are fragments with chromatogram data, the traces will reappear.

Understanding the Trace Display

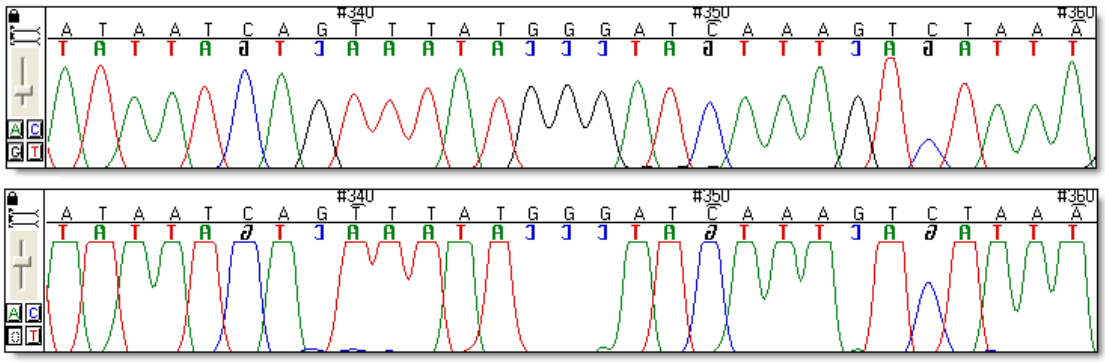
Above the traces are two lines of bases. The upper line displays the sequence as it currently appears in the editor, including any edits you may have made. The lower line displays the base calls as originally imported, including the trimmed data. The arrows to the left show the orientation and relative lengths of the trimmed and untrimmed data.

Below the arrows is a slider for changing the height of the trace. Below the slider are four buttons that let you hide the lanes for any or all of the traces. Click on the button to hide any lane. Click it again to display the lane. Figure 15-2 shows a contig with adjusted scales and hidden lanes.

Scrolling through a single chromatogram is simple. Position your cursor on the small rectangular slider button to the right of the trace. With the mouse button held down, you can move the slider down or up, depending on your current position within the sequence. Alternately, you can use the left arrow and right arrow keys to step through the sequence trace, base by base.

When you are working in a contig, if you highlight a base in the consensus line you can use the arrow keys to move through the consensus sequence. Note that as you do so all the contig chromatograms will move left or right, in step with the highlighted base.

Figure 15-2 Trace with hidden lanes and adjusted scales



You can specify the settings for trace display by going to the **Window** menu and choosing **User Preferences** and selecting **Chromatogram**. You can specify the height of the trace, the width from peak to peak, the way the lanes are identified and various screen display attributes. (See Chapter 21 on "Customizing Sequencher and User Preferences" for more information.)

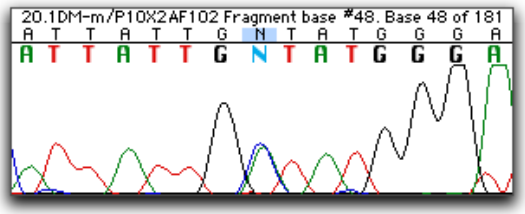
If you select the **Contig Chromatogram** section you can specify how many columns of traces you want to view at once, whether Sequencher should scroll to the selected column when you make a new selection in the contig and where the window should appear.

Note: If you hold down the **Ctrl** key while changing the scale, the new scale will apply to all the traces in that window. If you hold down the **Ctrl** key while clicking any of the buttons to hide a lane, the lane(s) will be hidden for all the traces in the window. (See Figure 15-2 above.)

Display secondary peak

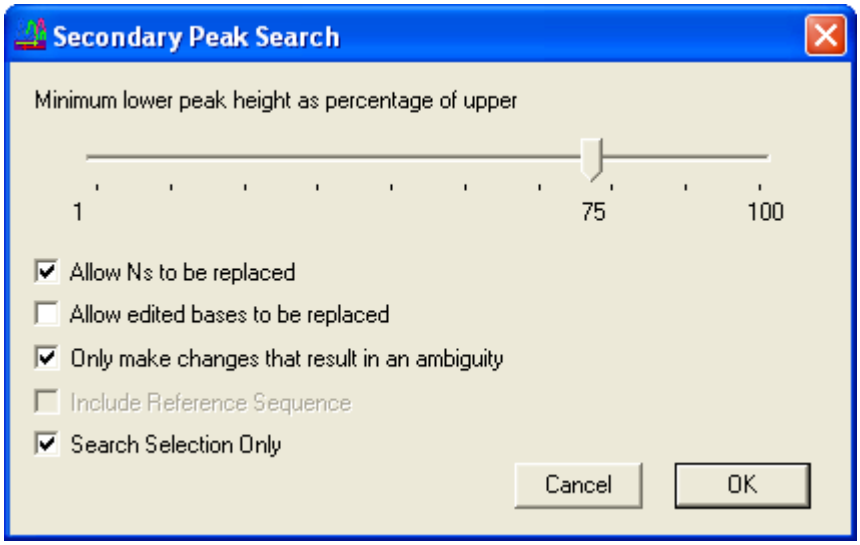
You can identify heterozygotes by identifying the second-highest peak beneath the primary peak. If you have a particular candidate for a heterozygote, first open the chromatogram for the individual sequence then click on the base call(s) as shown in Figure 15-3.

Figure 15-3 Trace with base selected



Go to the **Sequence** menu and choose **Call Secondary Peaks…**. A dialog box lets you specify how the secondary peak should be called (Figure 15-4).

Figure 15-4 Secondary peak parameters



The slider lets you specify how significant the second-highest peak must be to generate a change. In Figure 15-4 above, the slider is set so the secondary peak has to be at least 75% as high as the highest peak.

If you want Ns to be replaced then select the **Allow Ns to be replaced** checkbox.

If you've already edited bases by hand and don't want them changed, you must *deselect* the checkbox called **Allow edited bases to be replaced**.

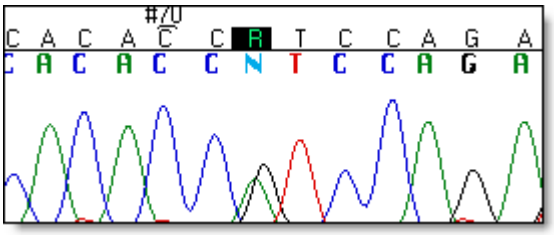
If you select **Only make changes that result in an ambiguity,** Sequencher may change an ambiguous base call to an A, C, G or T if it does not meet the secondary peak criteria. If you prefer it to remain ambiguous do not select this option.

If you wish to search just a range of bases, highlight them and check the **Search Selection Only** checkbox.

When you click **OK** to make the changes, a dialog box notes how many bases will be changed; you can choose to **Cancel** or **Continue** at that point. If you continue, the changes take effect. Once they have taken effect, unlike normal edits to a base, they cannot be undone.

Once you have pressed the **Continue** button, you can find all the changes made during the session by selecting the first base in the sequence (you can click on it right on the chromatogram) and going to the **Select** menu and clicking on **Next Edited Base**. After you execute any **Select Next** command, Sequencher creates a shortcut for that command in the spacebar so you can continue to search for your edits, using the same function over and over by simply pressing the space bar.

Figure 15-5 Base change generated by secondary peak



Select Next Commands

The **Select** menu provides you with several commands for navigating to bases which need further review or attention (see Table 15-1).

As you move your selection from each position that requires scrutiny to the next, you are likely to need to repeat these commands frequently. Therefore, Sequencher creates a shortcut for that command in the space bar after the first time you invoke a **Next** command. Press the space bar to move from one selected base to the next candidate.

To change any **Next** command to search in the reverse direction, combine the command with the **Shift** key.

If you use the **Next Met To Stop (>0bp)** you can highlight the next pair of start and stop codons in one of the three forward reading frames. The number of bases shown in this command depends on whether you specified a preference for a minimum length in the **User Preferences**.

Table 15-1 Summary of Select commands

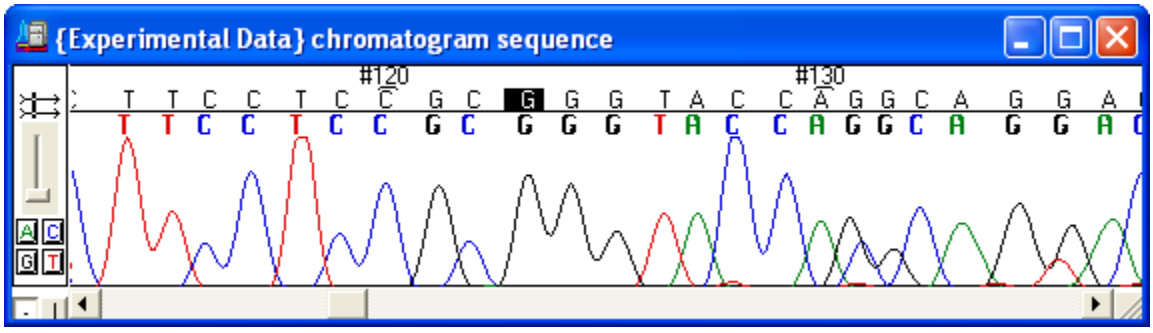
Select command	Contig
Next Ambiguous Base	Ambiguous bases, Disagreements, First base in large gap run.
Next Contig Disagree	Disagreements.
Next Edited Base	Edited bases.
Next Low Confidence Base	Any consensus base that is derived from a low confidence base.
Next Met To Stop (>0bp)	Any pair of Start and Stop codons in one of forward ORFs.

Editing bases from traces

Now that you can easily locate the bases that need attention, you can edit your data in either the contig editor window or the chromatogram display. As you edit, both windows update automatically. (Refer to Chapter 12 “Editing Contigs” for more information on editing.)

To edit a base while viewing a chromatogram, click on the current base call (upper line of text) you want to edit, then type the new base call (Figure 15-6). You can also select multiple bases for deletion by dragging the cursor across them.

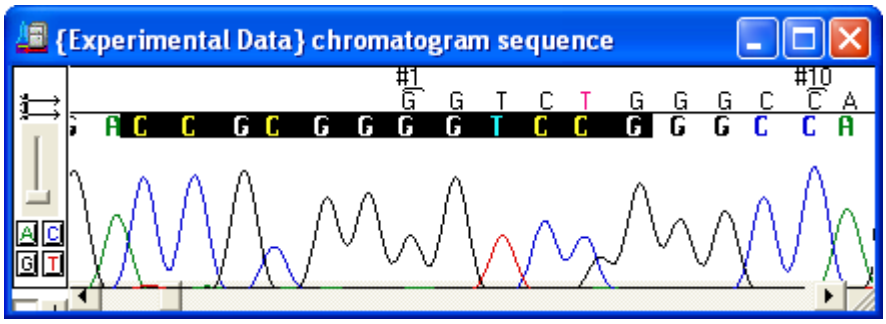
Figure 15-6 Clicking on a base displayed in a trace



Reverting to experimental data

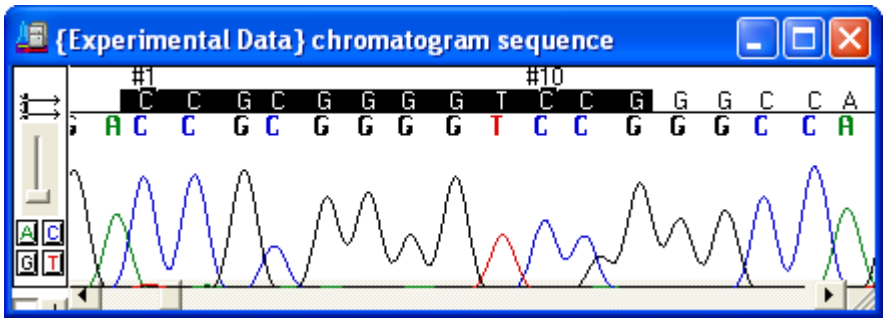
If you have deleted too much of a sequence or have made edits you want to remove, you can copy the original base calls back into the current sequence. From within the chromatogram, select the original bases (lower line of text) you want to restore (Figure 15-7).

Figure 15-7 Original base calls selected



Go to the **Sequence** menu and choose **Revert To Experimental Data.** Figure 15-8 shows the original bases restored.

Figure 15-8 Bases restored



Revert To Experimental Data is a powerful feature. It allows you to recover original data very precisely, as described above, or it can be used to globally revert an entire sequence or selection of sequences.

You can revert unassembled sequences to original data by selecting the sequences in the **Project Window** and executing the **Revert To Experimental Data** command.

Reverting assembled sequences back to experimental data

If you need to revert a sequence that has been assembled in a contig, select its icon from the left hand side of the Contig Editor. Go to the **Contig** menu and use the **Remove Selected Sequence**… command. Now you can use the **Revert To Experimental Data** command as described above.

If your sequence has trace data select all of the original bases. Go to the **Sequence** menu and choose **Revert To Experimental Data.**

Note: The **Extend Selection** command allows you to select all of the bases in a chromatogram without scrolling.

Adjusting trace positions

As the four dyes migrate in different lanes on certain automated sequencers, the synchronization among the four traces can deteriorate in output files. Sequencher allows for adjusting or "re-tracking" the position of one trace relative to the others.

When you are looking at a chromatogram, specify which of the four traces you want to shift by selecting a base in the *original* data line (the *lower* of the two lines of sequence characters).

Go to the **Edit** menu and choose **Shift** to display a submenu that lets you shift the base's line left or right by a few pixels. The adjustments made are to aid you in reviewing the original data; they will not be saved to the file when you save the project.

16. Restriction Maps

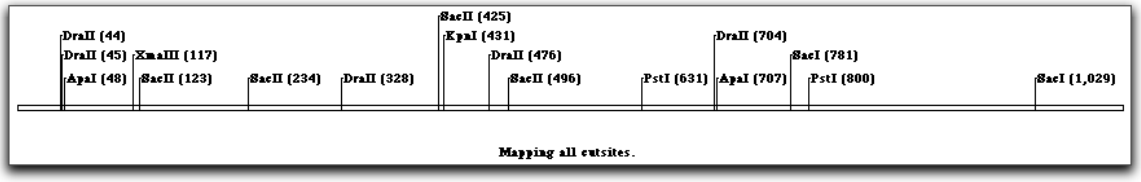
In this chapter we discuss restriction maps and how to work with them. We discuss choosing Restriction Map display options, selecting and changing selected enzymes, saving new or modified enzymes and copying the map.

Displaying a Restriction Map

You can view a Restriction Map for a sequence or a contig by selecting its icon in the Project Window and opening an editor. Go to the button bar of the editor and click on the **Cut Map** button. Alternately, from the editor, go the **View** menu and choose **Bases, Map, Overview…**, then choose **Restriction Map** from the submenu.

Figure 16-1 shows a restriction map display. The number in parentheses shows the cut location.

Figure 16-1 A single line restriction map



In this type of restriction map, clicking the name of an enzyme highlights all the locations where that enzyme cuts.

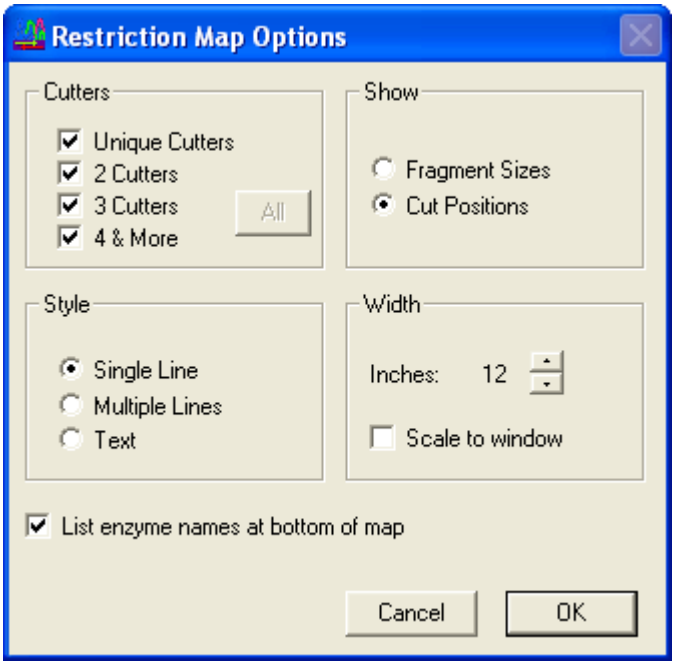
Restriction Map display options

How to select enzymes

To change the restriction map display, click the **Options** button in the button bar or go the **View** menu and choose **View Options…**

A dialog box (Figure 16-2) lets you change several attributes of the restriction map.

Figure 16-2 Restriction map options



Use the **Cutters:** checkboxes to display enzymes according to the number of times they would cut. For example, you could select just **2 Cutters**, that is, only those enzymes that cut in two locations. If you select the **All** button this turns on all of the checkboxes simultaneously.

How to show cut positions or fragment sizes

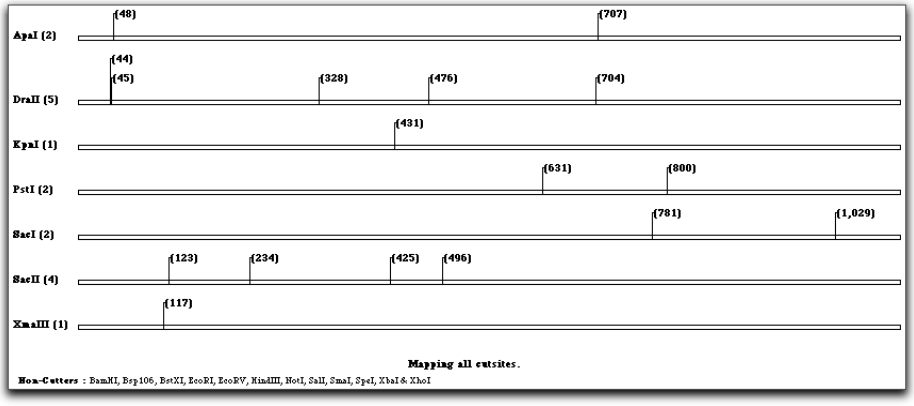
Use the **Show** options to choose the **Fragment Sizes** display or the **Cut Positions** display by clicking on one of the radio buttons.

How to set the map style

Use a radio button from the **Style:** pane to choose one of the three basic map styles. The **Single Line** map is the default display.

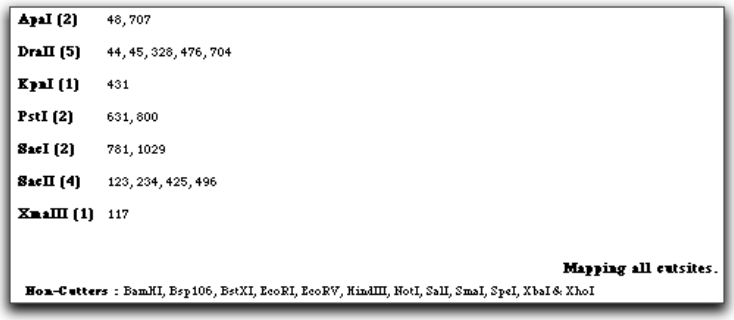
Choose **Multiple Line** map (Figure 16-3) if you want to show each enzyme on its own line.

Figure 16-3 Multiple line restriction map



Choose the **Text** representation (Figure 16-4) to list the cut sites in tabular form. In this view, the number in parentheses following each enzyme shows the number of times that particular enzyme cuts.

Figure 16-4 Text restriction map

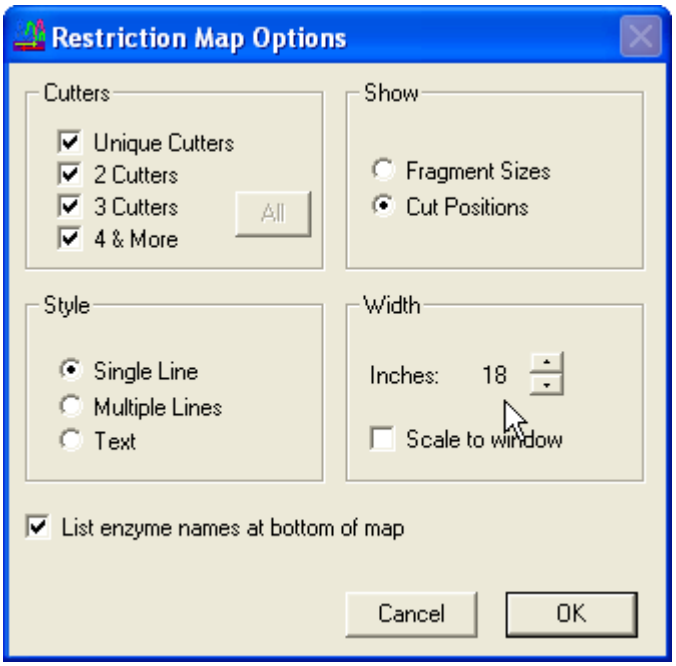

ApaI (2) 48, 707
DraII (5) 44, 45, 328, 476, 704
KpnI (1) 431
PstI (2) 631, 800
SacI (2) 781, 1029
SacII (4) 123, 234, 425, 496
XmaIII (1) 117
Mapping all cutsites.
Non-Cutters : BamHI, Bsp106, BstXI, EcoRI, EcoRV, HindIII, NotI, SalI, SmaI, SpeI, XbaI & XhoI

Setting the map width

The restriction map normally scales to the width of the window. When you print a restriction map, it scales to the size of the printed page.

To set the width of the diagram in inches, turn off the **Scale to window** checkbox and use the elevator button to set the width in inches (see Figure 16-5). This option is particularly useful when copying a map to paste into a drawing or presentation program.

Figure 16-5 Map options showing width in inches



Setting the map caption

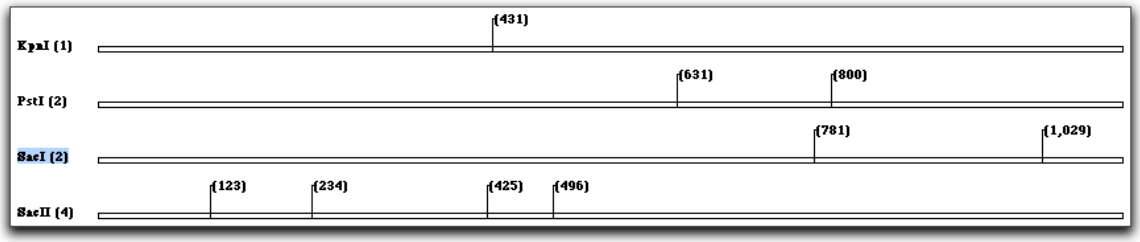
Use the checkbox **List enzymes names at bottom of map** to toggle the caption at the bottom of the restriction map. It lists all of the enzymes used to make the map (see Figure 16-5 above). The caption also lists as “Non-Cutters” enzymes that were not mapped. For

example, if you were displaying only unique cutters, then enzymes that cut more than once would be excluded. They would be listed as “Cutters that are not mapped.” Enzymes that did not cut at all would be listed as “Non-Cutters.”

Getting more information about an enzyme

You can get more information on the enzymes in your restriction map or edit the enzymes included in the map. In Figure 16-6, where the **Multiple Lines** style is being used, the SacI enzyme is selected.

Figure 16-6 SacI enzyme selected

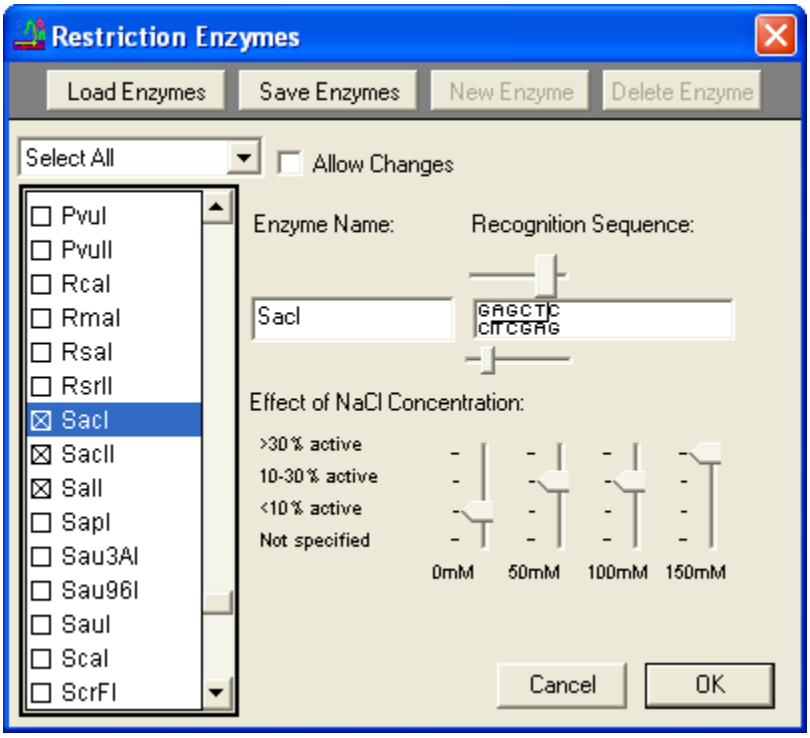


If you click on an enzyme label in the map and go the **File** menu and then choose **Get Info…** you can bring up the **Restriction Enzymes** dialog box. You can also use the keyboard shortcut **Ctrl+I.** Available enzymes appear in the scrolling list on the left side of the window. Sequencher highlights the enzyme you have selected and puts the appropriate information in the **Enzyme Name:** and **Recognition Sequence:** fields.

Changing the selected enzyme

You can also bring up the dialog by clicking the **Select Enzymes…** button in the button bar of any cut map. Alternately, you can go to the **Window** menu and choose **Specify Restriction Enzymes.** The enzyme editor dialog box is shown in Figure 16-7.

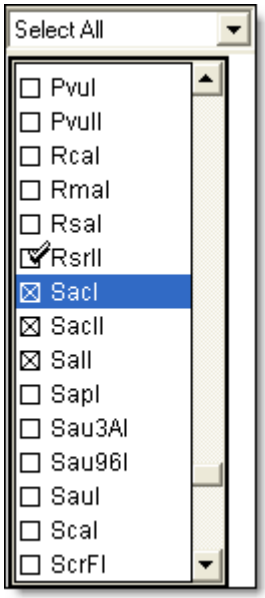
Figure 16-7 Restriction enzyme editor



Available enzymes appear in the scrolling list on the left side of the window. To view information about a particular enzyme, click on its name. Sequencher highlights the enzyme and puts the appropriate information in the **Enzyme Name** and **Recognition Sequence** fields.

If an enzyme is currently selected for display in the map, the square to the left of the name is filled with an 'x'. (An enzyme not selected for the map has an empty square.) When you move the cursor to the checkbox column, it changes to a check mark. To change your map selections, click the checkmarks to toggle them on or off. Figure 16-8 shows the list in detail.

Figure 16-8 Checkmarked enzyme list



If you are interested in enzymes with special characteristics, such as blunt ends or 3' overhangs, explore the commands in the **Select** menu within the dialog.

Start by choosing **Select None** to clear all selections. Each menu item you choose after that will add all of the enzymes with the specified characteristic. Selections are *cumulative*—no command (except **Select None**) removes any enzymes from the list. If, for example, you want all the enzymes that recognize sequences either 4 or 5 bases in length, choose **Select None** followed by **4-base** and then **5-base**.

When you are finished selecting enzymes, click **OK** in the lower right corner of the dialog box or press the **Return** key. Sequencher redraws your restriction map with the newly selected enzymes.

Changing the default enzymes

To change the default enzyme selections for items in a project, select the fragments or contigs you want to change. Go to the **Window** menu and choose **Specify Restriction Enzymes** while viewing the Project Window. Sequencher will warn you if no sequences are chosen. The enzymes you choose at this time are now the default set for any selected sequences or contigs in your project.

Adding and editing enzymes

Sequencher lets you add new enzymes to the list. Go to the top of the enzyme editor and select **Allow Changes**. All the sliders will now change from gray to blue. Click on **New Enzyme** to open a blank entry in the enzyme editor. Note that the **OK** button changes to **Done**.

You can now enter information in the **Enzyme Name** and **Recognition Sequence** fields. When you click **Done**, Sequencher adds the new enzyme to the list in alphabetical order.

Changing the recognition sequence

The recognition sequence field shows a double-stranded sequence. If you type in the sense strand, the anti-sense strand will be calculated automatically.

Sliders above and below the recognition sequence window are used to move the lines which indicate the cleavage positions. You can use the mouse to move the sliders left and right.

Note: The maximum length for a recognition sequence is fifteen bases. If you exceed this length, the right end will be truncated.

Salt concentration effects

You can record buffer sensitivity with the **Effect of NaCl Concentration:** slider controls. The default is an unspecified value for all concentrations of NaCl. Use the slider controls to indicate the percentage activity for a given NaCl concentration.

Saving new or modified enzymes

If you enter an enzyme that you will want to use later, click **Save Enzymes** to save the enzyme set as a file. Click **Load Enzymes** to retrieve a saved file.

Setting the sequence selection in a Sequence Editor

When you click on the **Cut Map** button, a full-page cut-map is displayed. If you click between two cut sites, Sequencher highlights the resulting restriction fragment. Hold down the **Shift** key and click in the rectangle between other cut sites. Your selection is extended from the original site you selected. When you toggle back to the **Bases** view you will see that the sequence representing the chosen region also has been highlighted.

If you wish to display the sequence and the cut map at the same time, click the button showing the staggered enzyme cut icon. A window with the restriction map appears. You can adjust the size of the panes by dragging the splitter bar that separates the scroll bars for each pane.

Like the full-page cut map, this window is linked to the Sequence Editor. If you click on the name of an enzyme, Sequencher boxes the recognition sites. If you click between two cut sites, Sequencher highlights the resulting restriction fragment.

Hold down the **Shift** key and click in the rectangle between other cut sites. Your selection is extended from the original site you selected. You can copy the bases between the two cut sites by going to the **Edit** menu and choosing **Copy Selection**.

You can use this command to excise a particular restriction fragment by clicking on it in the cut map and switching back to the **Bases** view. Once you are in the **Bases** view, you can remove the highlighted fragment by going to the **Edit** menu and using the **Cut Selection** command.

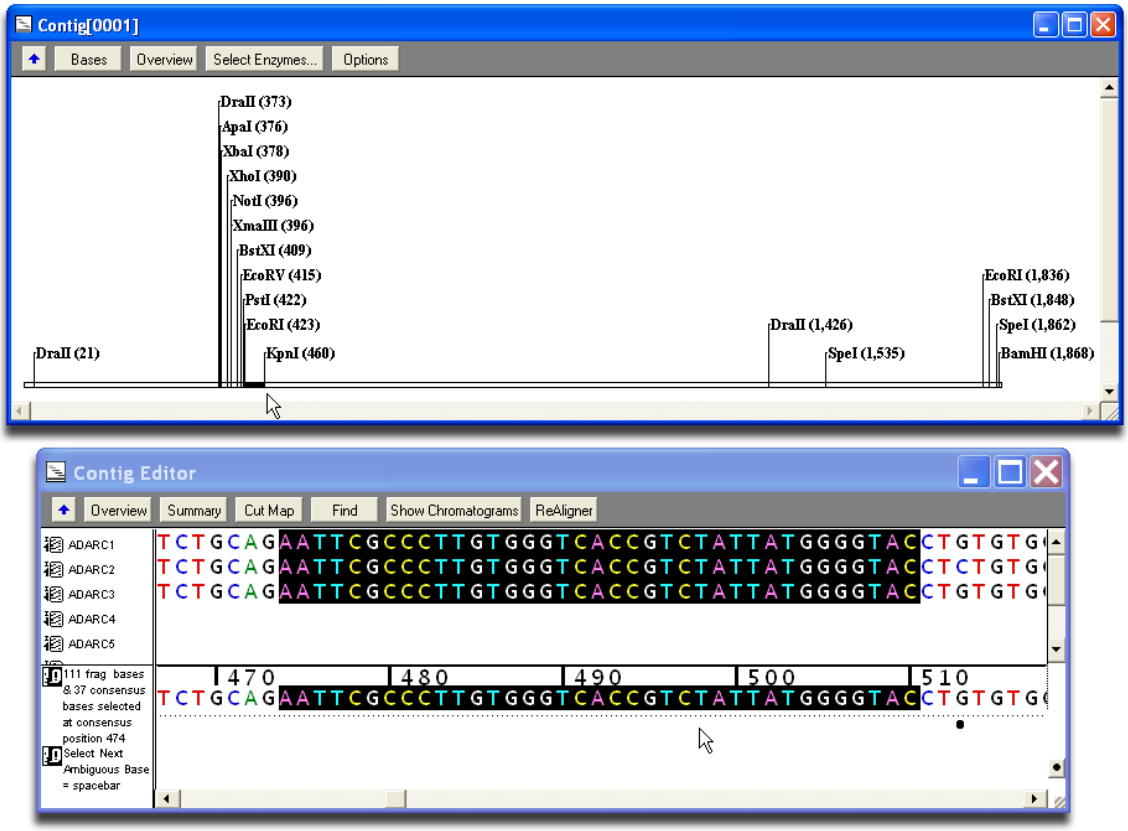
Setting the sequence selection in a contig editor

When you are working in the **Bases** view of a contig editor, you can display a restriction enzyme map by clicking on the **Cut Map** button in the editor's button bar. You can also

display this view by going to the **View** menu, selecting **Bases, Map, Overview...** and then choosing **Restriction Map** from the submenu.

If you click between two cut sites, Sequencher highlights the resulting restriction fragment. Hold down the **Shift** key and click in the rectangle between other cut sites. Your selection is extended from the original site you selected. When you return to the **Bases** view you will see that the sequence representing the chosen region has been highlighted also. (See Figure 16-9.).

Figure 16-9 Fragment selected in map view and its equivalent bases highlighted



Copying the map

After you have made a selection in the restriction map, as in Figure 16-9 above, you can copy the bases between the two cut sites by going to the **Edit** menu and choosing **Copy Selection**.

You can also copy the entire diagram for use in an illustration program by going to the **Edit** menu and choosing **Copy As** and **Picture** from the submenu.

17. Motifs and Features

In this chapter we describe Sequencher's ability to locate and highlight subsequences that you specify as motifs. You will learn how to create, edit and display features and list a sequence's features.

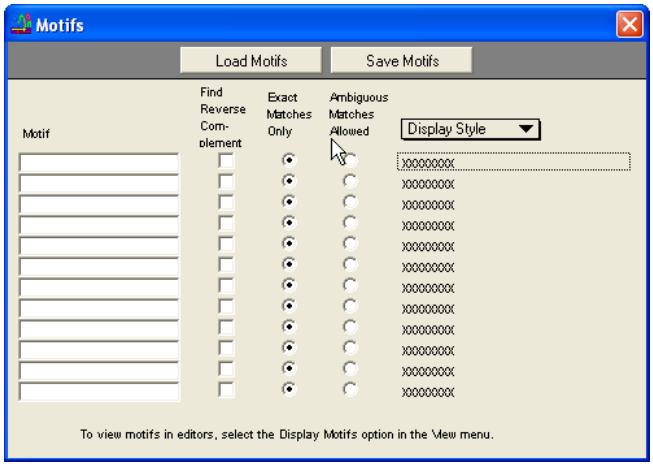
Motifs

Sequencher lets you define up to 12 patterns of 50 or fewer bases as motifs.

Entering motifs

You define new motifs by going to the **Window** menu and clicking on the **Motif Definitions…** command.

Figure 17-1 Motif entry window



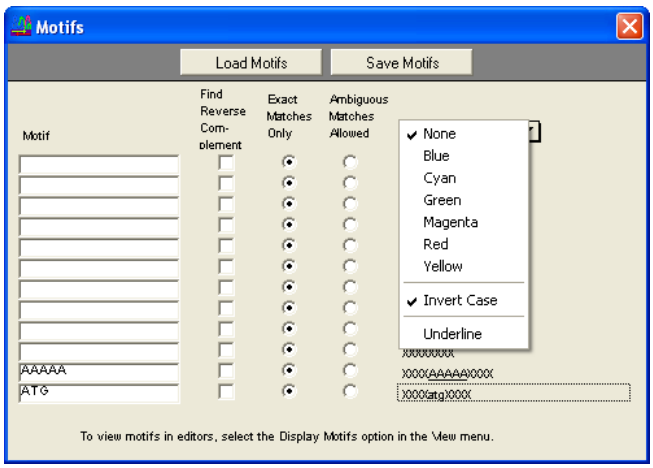
The default setting shows exact matches only. To show the reverse complement for a motif, click the **Find Reverse Complement** checkbox. Use the radio buttons to specify

whether you want to highlight the motif using the exact bases you entered or using ambiguous matches.

To specify how you would like the motif you are entering to be highlighted, click on the **Display Style** pull down menu (Figure 17-2) to choose color, case and underlining while the cursor is still blinking in the entry field.

In the example (Figure 17-2), every instance of 'AAAAA' will be underlined and every start codon (ATG) will be shown in lower-case letters. You can save the motifs in a file on your hard disk by clicking on the **Save Motifs** button. Enter a suitable name for your motif file in the **Save** dialog window.

Figure 17-2 Display style pull down menu



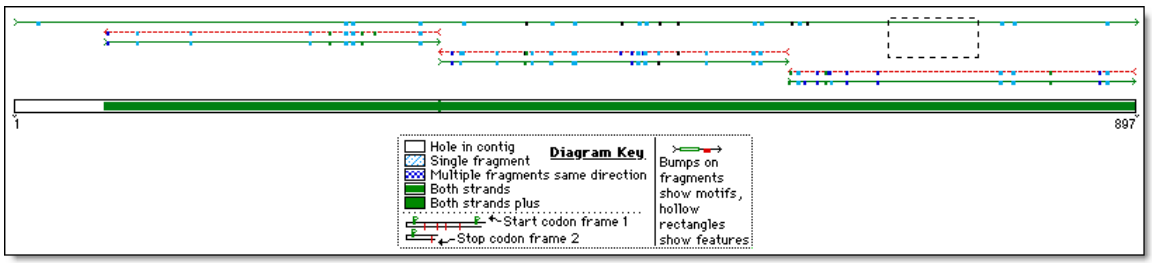
Quick motif entry

Select the bases in a Sequence Editor, then go to the **Edit** menu and click on **Enter Selection As Motif.** If you choose this command the default style will be used (see Chapter 21 "Customizing Sequencher and User Preferences").

Displaying motifs

To use the motifs you have entered, go the **View** menu and choose **Display Motifs.** Figure 17-3 shows how motifs are displayed in the contig overview: they are the small downward facing colored blocks on the arrows representing the sequences. The colors are those you set in the **Motif Default Style** menu (see Chapter 21 "Customizing Sequencher and User Preferences").

Figure 17-3 Overview showing motifs



In the **Bases** view of a sequence motifs will appear highlighted in the manner you specified. You may need to deselect **Display Color Bases, Colors As Backgrounds** or **Display Base Confidences** to see the **Motifs** more clearly. In Figure 17-4, for example, all start codons are in lower case type and blue with only **Colors as Backgrounds** on.

Figure 17-4 Sequence Editor showing motifs

If you have a certain set of motifs you want to use for several projects (for example, if you use motif definitions to highlight certain restriction sites), you can save your definitions to a file. To do so, click **Save Motifs** at the top of the motifs window and follow the instructions. Then click the **Load Motifs** button to use those definitions in other projects.

Features

What is a feature

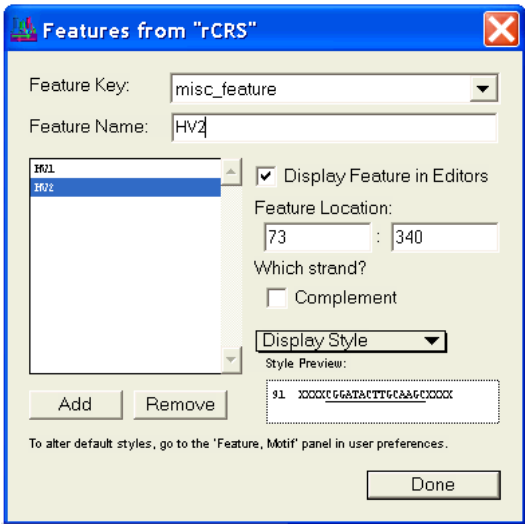
A feature represents a subsequence or region of a sequence which has some biological significance. Assigning a feature is a method of describing and naming such a region. Sequencher supports both personal features and GenBank standardized features. The Feature Editor enables you to name and highlight sequence regions of special interest.

Note: Features will not appear unless you go the **View** menu and select **Display Features**.

Creating and editing features

To create or edit features in a sequence, click on the sequence icon. Go to the **Sequence** menu and choose **Edit Features…** to display the **Feature Editor** (Figure 17-5).

Figure 17-5 Feature Editor dialog box



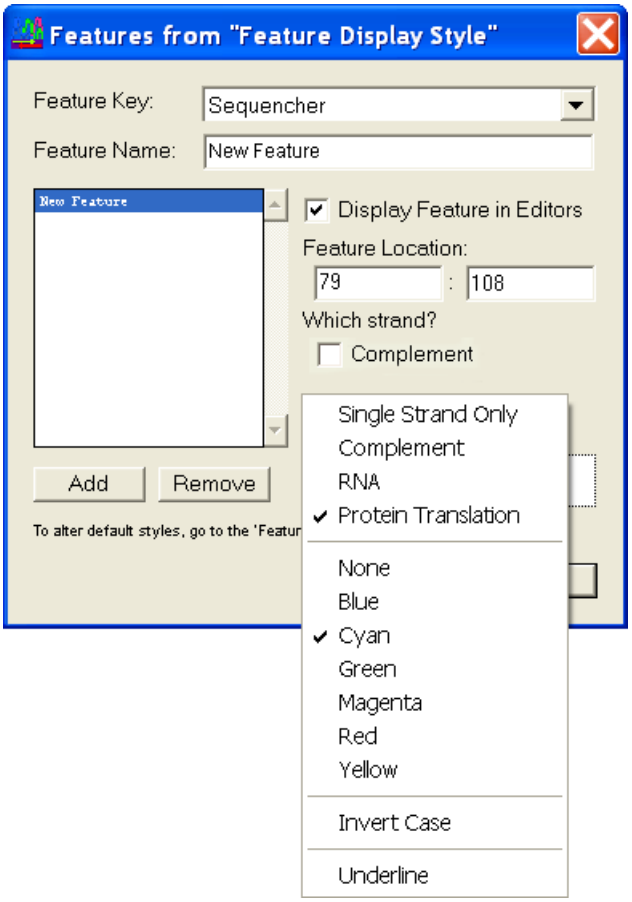
To create a new feature, click the **Add** button in the dialog box. From the **Feature Key:** menu choose **Sequencher** if you are creating personal annotations, or choose one of the GenBank keys.

If you have chosen **Sequencher** as your **Feature Key:**, type a name in the **Feature Name:** field. Enter the starting and ending base numbers for the feature in the **Feature Location** boxes. The default strand is **5'-> 3'** but you can change this using the **Complement** check box. Use the **Display Style** pull down menu (Figure 17-6) to specify strand type, color, case and underlining. You may select one attribute from each of the four style categories. When you have finished entering or editing your features, click **Done.**

If you have chosen a GenBank **Feature Key:**, Sequencher will insert a default name in the **Feature Name:** field using the Feature Key and Feature Qualifier. You may edit this if you wish. Enter the starting and ending base numbers for the feature in the **Feature Location** boxes. The default strand is **5'-> 3'**but you can change this using the **Complement** check box. Use the **Display Style** pull down menu (Figure 17-6) to specify strand type, color, case and underlining. You may select one attribute from each of the four style categories. When you have finished entering or editing your features, click **Done.**

Check the **Display Feature in Editors** checkbox and then go to the **View** menu and click on **Display Features** to have your feature appear in the sequence or contig editor and the associated overviews. When you set a feature to include **Protein** or **DNA-RNA** this will only be displayed in appropriate views.

Figure 17-6 The display style menu



Editing an existing feature

To edit an existing feature, go to the **Sequence** menu and choose **Edit Features…** To display the **Feature Editor,** click on the name of the feature in the feature list. Use the **Display Style** pull down menu to specify new display attributes. You can also change the **Feature Key:**, edit the **Feature Name:** and the **Feature Location:**. For example, you could change a Sequencher annotation into a GenBank feature.

If you want to remove a feature, select its name from the list and click on the **Remove** button.

When you have finished entering or editing your features, click **Done.** The features will appear in the Sequence Editor if you have checked the **Display Features in Editors** checkbox and **Display Features** in the **View** menu.

Quick feature creation

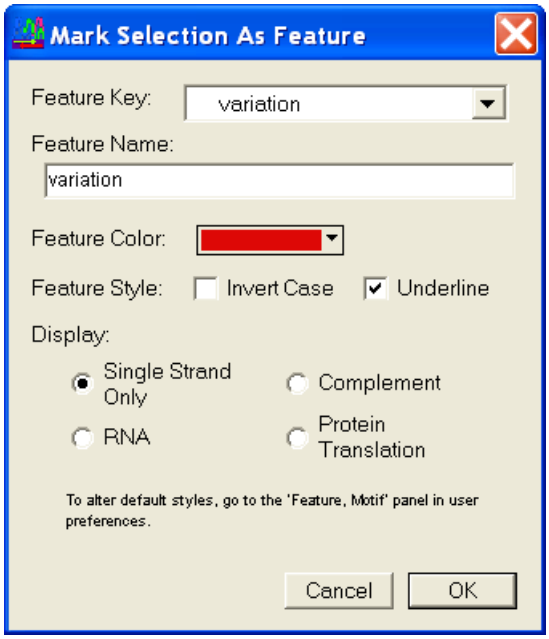
To create a feature quickly, select a region in the sequence or contig editor and then go to the **Sequence** menu and choose **Mark Selection As Feature.** Sequencher displays a dialog box called **Mark Selection As Feature** that lets you assign a Feature Key, name the feature and specify display attributes (Figure 17-7). When you have finished assigning the feature attributes click the **OK** button.

If you create a feature while selecting bases in the consensus of a contig, you will actually create a feature for each of the sequences that contribute to the consensus at the selected bases.

Note: If a fragment spans only a portion of the bases selected for a feature that fragment will not carry the new feature.

If you do not define display properties, the display attributes will default to the ones you specified in the **Feature, Motif** section of the **User Preference** settings. (See Chapter 21 "Customizing Sequencher and User Preferences" for more information.)

Figure 17-7 The Mark Selection As Feature dialog box



Showing and hiding features

Go to the **View** menu and use **Display Features** to toggle feature display on and off. For each individual feature you must also enable or disable its display in User Preferences. (See Chapter 21 "Customizing Sequencher and User Preferences" for more information.)

Listing a sequence's features

If you want to get a complete list of all your sequence's features, go to the **Sequence** menu and choose **Feature Listing**. This will display a text list of each feature giving its name, Feature Key, range, style and any Feature Qualifiers. You can print out this listing for your records.

Note: If you have imported a GenBank sequence and Sequencher does not recognize any Feature Keys this list may differ.

18. Finding items

In this chapter you will learn how to find items according to specific criteria. It also discusses how to refine your subsequence search and search for and find particular items in the Project Window. This chapter also discusses how to find open windows and editors.

Finding the Project Window

You can always bring the Project Window to the front by going to the **Window** menu and choosing **Project Window**.

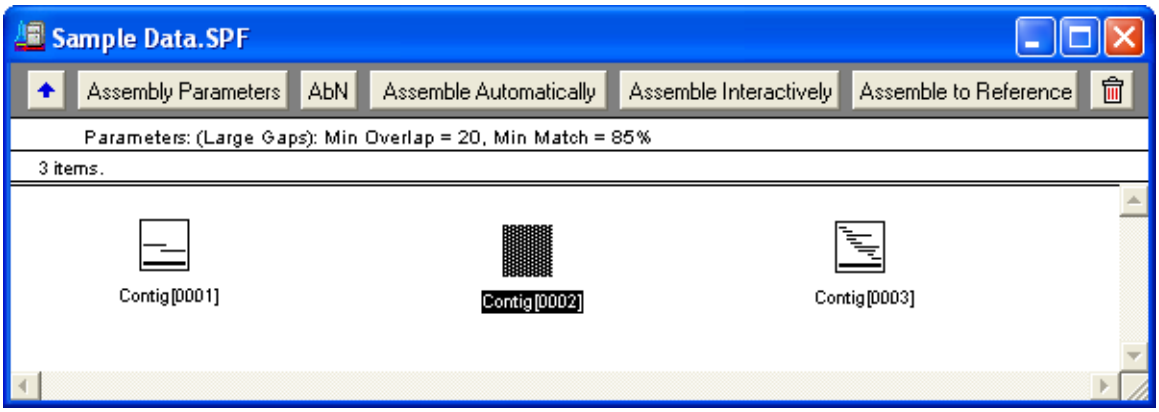
Finding open windows with menu commands

Sequencher has a number of commands in the **Window** menu for finding open windows hidden behind other windows—**Project Window**, **Variance Table, Translated Variance Table**, **Chromatograms, Contig Editors,** and **Sequence Editors**. Choose the appropriate command to bring the window you want to the front of your screen. If one or another of the commands is dimmed, you do not currently have that editor open. If you have several windows of a type open, these will be listed as submenus of that type.

Finding open windows in the Project View

If you are looking at a Project Window, any icons that already have open editors are dimmed. To move an open editor to the front, just double-click the dimmed icon, for example, the icon for Contig[0002] shown in Figure 18-1.

Figure 18-1 A dimmed icon



Finding the current selection

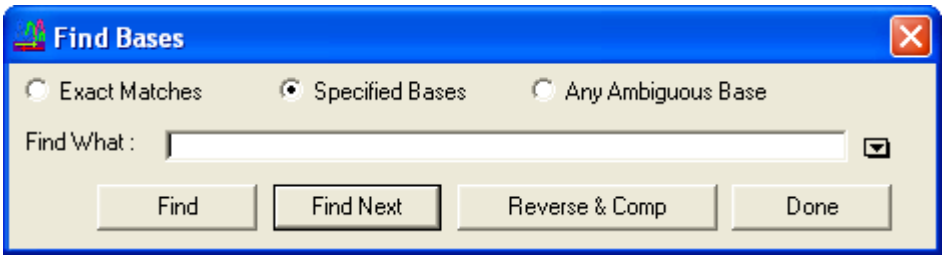
In a sequence or contig editor, you can always scroll to the currently selected bases by using the **Enter** key on the number keypad.

Searching in Open Windows

Finding a subsequence

You can search for a specific subsequence in an open sequence or contig editor. Go to the **Select** menu and choose **Find Bases...**; Sequencher displays the **Find** dialog box shown in Figure 18-2. In the **Find What:** field, type the string of bases you want to find. Use the drop down menu to specify the recognition sequence of one of a subset of restriction enzymes available. Use the **Other** option in the drop down menu to type in the name of your enzyme of interest. If it is in the database, the recognition sequence will be placed in the **Find What:** box.

Figure 18-2 Find Bases dialog box



Find locates the first occurrence of the search string anywhere and then selects the string in the editor. **Find Next** finds the next occurrence after the current selection, whether it is a match found by Sequencher or a selection that you marked.

The **Find Bases** dialog box stays in front until you click **Done**. If you want to repeat the same subsequence search but without invoking the **Find Bases** dialog box you can go to the **Select** menu and use the **Find Bases Again** command.

Refining a subsequence search

You can refine any of your searches by using one of the radio buttons. The default setting is **Exact Matches.**

Exact matching

If you click **Exact Matches** the search is based on exact character matches and disregards matches that occur due to ambiguities. For example, AGGTW would only match AGGTW.

Matching ambiguous bases

If you click **Specified Bases** the search finds only matches that are identical or more specific than the sequence you entered in the **Find What:** field. For example, the string AGGTW would match AGGTT or AGGTA, since W is the ambiguity code for T *or* A.

Any ambiguous match

If you click **Any Ambiguous Match** the search will find any degenerate DNA that may match the sequence you have entered, including matches consisting of all unknown bases. For example, AGGTW would match NNGTN or even NNNNN.

Finding bases by number

To select a range of bases quickly, go to the **Select** menu and choose **Bases by Number**. A dialog box lets you enter the range of bases you want to highlight.

Extend Selection

You can extend the current selection to the beginning or end of your sequence by going to the **Select** menu and choosing **Extend Selection.** Then choose the suboption to **Left End** or **To Right End.**

Note: In a contig chromatogram the **Extend Selection** command can extend your selection across the entire sequence in one click.

Other search options

You can also search for ambiguous bases by going to the **Select** menu and using the **Next Ambiguous Base** command. You can search for edited bases using the **Next Edited Base** option. (These are described in more detail in Chapter 8 “The Sequence Editor”.) You can also search for potential open reading frames using the **Next Met to Stop (> 0b)** option. (This is described in more detail in Chapter 21 “Customizing Sequencher and User Preferences.”)

Project Window selection

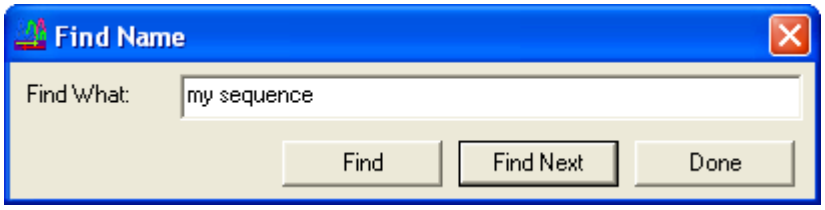
Finding an item by name

To locate a specific sequence or contig, select the Project Window. Type the name of the item you are trying to find, for example, ‘Sequence C.’ As you type, Sequencher starts looking in the Project Window for an icon with the name you are typing and highlights a match. If it finds more than one, it highlights the first one.

Note: If you pause between characters for more than a few seconds, the program assumes you've started typing a different name and looks for the new name.

Another way to locate a sequence is to go to the **Select** menu and choose **Item Named....** This brings up the dialog box shown in Figure 18-3.

Figure 18-3 Item Named… dialog box



Type the name of the sequence you want to find and click **Find**.

Sequencher searches through all the sequences displayed in the Project Window for a name that matches the name you entered.

- When it finds a sequence, the sequence is displayed and highlighted in the Project Window.
- If Sequencher does not find a sequence with the exact name you typed, it looks for a sequence with a name starting with the first few letters you typed.
- If Sequencher still does not find a match, it looks at all the sequences contained in contigs. If it finds a matching sequence within a contig, it highlights the contig that contains the matching sequence.
- If Sequencher can find no matches anywhere, it stops searching and informs you that the item was not found.
- If there is more than one sequence with a matching name, Sequencher selects and highlights the first one. If that is not the one you want, click **Find Next**. This continues the search in alphabetical order.

Note: The **Find** window is movable.

Selecting all items that…

The **Select** menu's **All Items That** command provides a series of suboptions based on different criteria. For example, you can choose sequences that have chromatograms or that include a particular annotation in the comments field. The menus are context sensitive so not all options may be available for a particular view.

Most options will have a dialog box containing a number of important buttons, which control the results of your search. The **Select All Matches** button will select matches based on your search criteria. If you change your criteria and search again, these results will be added to your initial search. If you don't want this information added, you must choose **Select None.**

A status line will indicate how many items matched out of the total number of items available. You can dismiss the window by clicking on **Done** or by simply clicking on the Project Window or performing another action.

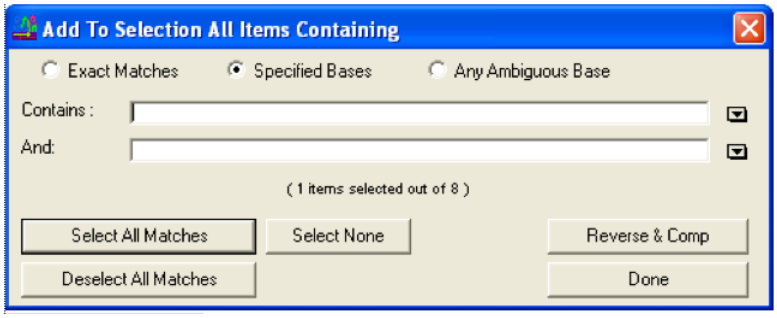
If you want to select all the items that *do not* contain the specified string, go to **Select All Items That,** click on **All Items That** command, and click the **Done** button. Then go to the **Select** menu and choose **Invert Selection.**

Contain a subsequence

To select all the sequences and contigs in the Project Window that contain a specific string of bases, you can use the **Contain Subsequence** command. Go to the **Select** menu, click on the **All Items That** command and then choose the **Contain Subsequence** menu item to display a dialog box (Figure 18-4).

Note: You can do a Boolean 'and' search using two specifiers for the subsequence. Enter the first specifier in the **Contains:** field and the second in the **And:** field. There are also two drop down menus to assist you in the search for specific restriction enzyme sites.

Figure 18-4 Contain subsequence… dialog box



Contain Items Named

If you want to find which contig contains a specific fragment, go to the **Select** menu and choose **All Items That** and then **Contain Items Named…** to display a dialog box (Figure 18-5). This lets you select items in the Project Window that contain a fragment bearing the specified name.

Figure 18-5 Contain Items Named… dialog box

Contain Items With Names Containing

If you are looking for several fragments with related or similar names and need to check which contig or refrigerator they are in, go to the **Select** menu and choose **All Items That.** Click on the submenu option **Contain Items With Names Containing…** to display

a dialog box (Figure 18-6). This lets you select items in the Project Window where the search string is based on only part of the complete name.

Figure 18-6 Contain Items With Names Containing… dialog box

Have Chromatograms

If you want to find out which fragments have a chromatogram or which contigs contain fragments with chromatograms, go to the **Select** menu and choose **All Items That** and then **Have Chromatograms.**

Have Comments Containing…

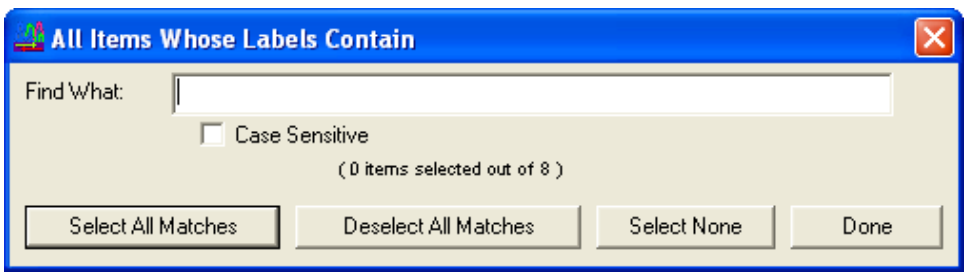
If you want to select all the sequences in the Project Window whose comments contain a specific text string, go to the Select menu and choose **All Items That** and then **Have Comments Containing…** to display a dialog box (Figure 18-7).

Figure 18-7 Have Comments Containing… dialog box

Have Labels Containing…

If you want to select all the sequences in the Project Window whose label contains a specific text string, go to the **Select** menu and choose **All Items That** and then **Have Labels Containing…** to display a dialog box (Figure 18-8).

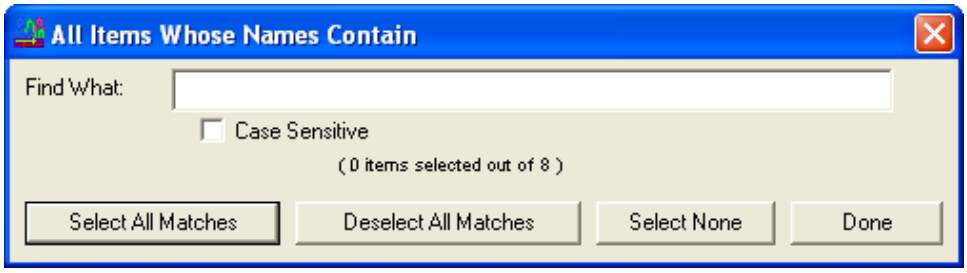
Figure 18-8 Have Labels Containing… dialog box



Have Names Containing...

If you want to select all the sequences in the Project Window whose names contain a specific text string, go to the **Select** menu and choose **All Items That** and then **Have Names Containing...** to display a dialog box (Figure 18-9).

Figure 18-9 Have Names Containing...dialog box



Were Edited On Or Since...

If you want to select all the sequences and contigs in the Project Window that were edited on a specified day, go to the **Select** menu and choose **All Items That** and then **Were Edited On Or Since...** to display a dialog box. You can use the drop down menu to specify a preset day such as today, yesterday, 1 week ago or even 2 months ago. You can also type a date in the **Find What:** box.

Note: You can also include items edited on all the days since a specific date by checking the **And All Days Since** box.

19. Exporting Data

In this chapter you will learn how to export your work. This chapter discusses how to export sequences, consensuses and contigs, and also how to export defined data such as selected bases and subprojects.

Exporting data from Sequencher

Sequencher has a rich set of export formats. You can export sequence(s), contig(s) or consensus sequence(s), subsets of your data, the Variance Table, the Summary Report and even protein translations.

Sequencher offers you a number of options depending on the type of data you choose to export. For each data type you can select a choice of format and the location for the export. For some data types there are additional options. Sequencher may offer a default for the export. Unless you deselect the option to use default names, Sequencher saves the file(s) in the default name. The **Export** dialog box will default to the last location and choice of format used.

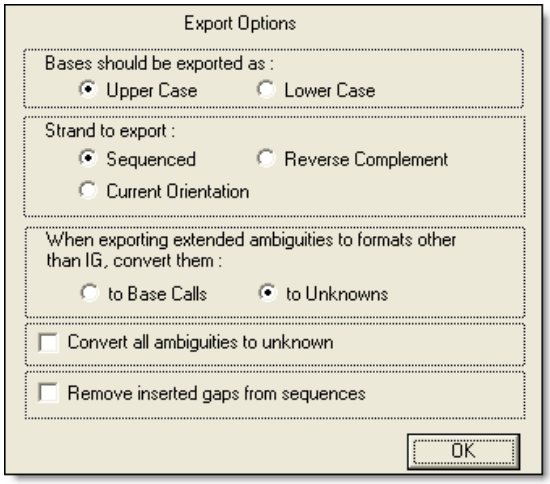
All of the export commands are accessible from the **File** menu.

Exporting Sequences

To export a sequence or a number of sequences, click on the item(s) you want to export. Go to the **File** menu and click on **Export** and choose **Sequence(s)...** from the submenu.

Sequencher provides a context sensitive dialog box for changing the export options. (See Figure 19-1.)

Figure 19-1 Export Options dialog box

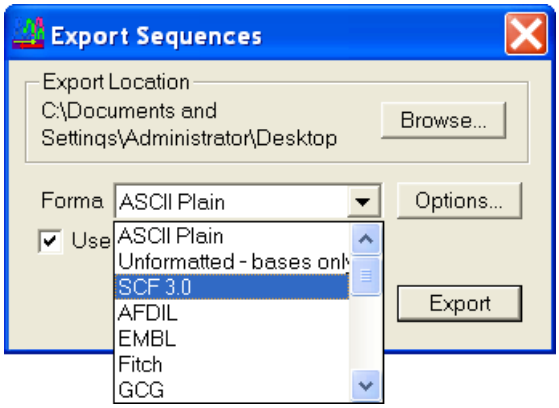


The **Browse…** button opens a window called **Choose a Folder.** From this window you can choose where to store the new text file(s). You may create a new folder in which to store your sequences by clicking on the **New Folder** button at the bottom left of the window. Click on the **Choose** button to record your chosen location.

The **Format:** drop down menu allows you to choose a format type for your exported sequence. You may set the case of the exported sequence, the sequence strand to export, and the treatment of ambiguities and gaps. To choose these settings click on the **Options…** button, select the options you want and click **OK** (see Figure 19-2).

Finally, click on the **Export** button to execute the command.

Figure 19-2 Export Sequences drop down menu



Note: If you have added features to a sequence and chosen GenBank as your export format, these will be exported as a GenBank Feature Table.

Exporting a Consensus

To export a consensus sequence, click on the contig icon(s) in the Project Window. Then go to the **File** menu and choose **Export** and **Consensus…** from the submenu. The

Format: drop down menu allows you to choose the format type for your exported sequence. Clicking on the **Options…** button displays the window where you turn options on or off. In most cases you will want to just click on **Export** and use the original consensus name.

The **Browse…** button opens a window called **Browse For Folder.** From this window you can choose where to store the new text file(s). You may create a new folder in which to store your sequences by clicking on the **Make New Folder** button at the bottom left of the window. Click on the **OK** button to record your chosen location.

Uncheck the **Use Default Names** checkbox if you wish to set a specific export name.

Exporting Contigs

To export a contig, click on the item you want to export. Go to the **File** menu and click on **Export** and **Contig…** from the submenu.

Sequencher displays the export dialog. The browse button allows you to choose a disk and folder to store the new text file(s). The format button allows you to choose the format type for your exported contig. You can choose from MSF, Nexus Interleaved, Nexus Sequential, Aligned FastA and CAF. You may also set a number of options. Clicking on the **Options…** button displays the window where you turn options on or off. Finally, clicking on the **Export** button executes the command.

The **Use Default Names** option gives each export the name of its contig. Uncheck the **Use Default Names** checkbox if you wish to set a specific export name.

The **Browse…** button opens a window called **Browse For Folder.** From this window you can choose where to store the new text file(s). You may create a new folder in which to store your sequences by clicking on the **Make New Folder** button at the bottom left of the window. Click on the **OK** button to record your chosen location.

If you execute the **Export** command with the **Sequence…** suboption while a contig is selected, Sequencher will create a folder with the name of the contig. The folder contains all of the sequences in the contig.

Exporting Selection as subproject

If you wish to share your data with colleagues who may have older versions of Sequencher or another program, choose **Export** from the file menu and then **Selection As Subproject…,** then select the appropriate version or type from the **Format** dropdown menu.

In certain cases, you may want to move some but not all items from one Sequencher project into another. Working in the Project Window, select the sequence fragments and contigs you want to export. Go to the **File** menu and choose **Export** and then choose **Selection As Subproject.** Sequencher creates a new project containing only the items you selected. Now you can import the newly created sub-project into the chosen Sequencher project.

You should remember to provide a relevant name to define the new project, because you have derived it from a selection of multiple items.

Note: Unlike other **Export** commands which may create a number of files on your hard disk, this command will create only one new project.

Export Selected Bases

You may want to export selected bases rather than an entire sequence. To do so, drag the mouse to highlight your selected bases or go to the **Select** menu and use the **Bases by Number** command. Then go to the **File** menu and choose **Export** and click on **Selected Bases…** from the submenu. You will then see the **Save As** window, which allows you to choose your format options and preferred location.

Export Options

If you click the **Options** button in an export dialog box, Sequencher provides a context sensitive dialog box for changing the export options. Select the options you want and click **Export**.

Export Formats

When you export a file, you may want to specify a file format so other software can access it. Click the **Format:** dropdown menu in the export dialog box. The menu will display a list of the available formats.

The Sequencher **Export** menu is context sensitive, which means that it varies based on whether your selection contains a sequence, a contig or both. For each kind of export, you will be prompted to apply appropriate formats. See Table below for more information.

Table 19-1 Examples of Sequencher's Export formats

	Export format type	Description
Sequence & Consensus formats	ASCII plain, Unformatted,	Plain text
	AFDIL, Fitch, Genentech, IG, Strider	Specialist and legacy formats
	GenBank, NBRF, EMBL	Database formats
	FASTA, FASTA concatenated GCG	Commonly used formats
	Nexus/PAUP interleaved, Nexus/PAUP sequential, Phylip, Phylip 3, Phylip4	Phylogenetic program format
	SCF 2.0, SCF 3.0	Standard Chromatogram Format
Contig formats	CAF	Common Assembly Format
	MSF	Multiple Sequence Format
	FASTA aligned	Phylogenetic program format
	Nexus/Paup	Phylogenetic program format
Subproject formats	CAF	Common Assembly Format
	FASTA	Commonly used formats
	Sequencher	Older project file formats

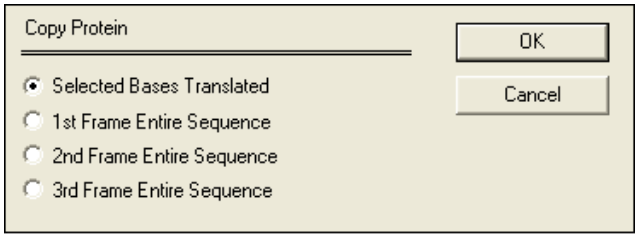
Note: SCF - Standard Chromatogram File format is a file format for exchanging data between different sequencing systems that use four color chromatograms to show sequencing data. Most current sequencing hardware supports this file standard.

Exporting protein translations

Sequencher can automatically translate bases into proteins as it copies items to the clipboard. These protein sequences can then be exported to another file or program.

To perform this, select a portion or all of the sequence and then go to the **Edit** menu and choose **Copy As** and **Protein Translation…** from the submenu. Sequencher asks you whether you want to translate only the selected bases or whether you want the entire sequence but in a specific frame. The default setting is to translate only the bases you selected. You can then take this copy and paste it into another program as you would any other information.

Figure 19-3 Copy As Protein Translation options dialog box



Alternatively, you can also go to the **File** menu. Click on the **Export** command and choose **Selected Bases as Protein…** from the submenu. You will be presented with a window similar to that above (see Figure 19-3). You will then be prompted with a **Save As:** window.

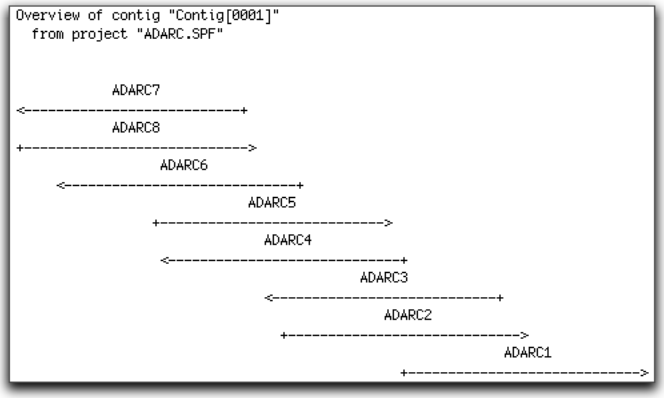
Export Overviews

If you are looking at the schematic view of a contig, you can generate a text version map of the fragments, as shown in Figure 19-4, for viewing in other programs or word processors. There are two ways to do this.

1. Go to the **File** menu and choose **Export**, then **Overview as Text** from the submenu. Sequencher will treat the map as any other exported file.
2. Alternatively, you can copy the overview as a picture and paste it into other programs. Select the overview window by clicking on it, then go to the **Edit** window and choose **Copy As** and **Picture** from the submenu.

Note: When viewing the output, you must specify a fixed-width font to preserve the scale of the arrows in relation to each other and the fragment names.

Figure 19-4 ASCII map of sequence



Exporting contig summaries

When you are displaying the **Summary** view of a contig (see “Summary Report View” in Chapter 12, “Editing Contigs”), you can export the contents of the report to a text file. Go to the **File** menu and choose **Export** then the **Summary…** submenu. This is particularly useful if you wish to add annotations to a report of your assembly in standard page format.

20. Output

In this chapter we discuss printing options. We will explain how to set up pages, copy pictures, print and use different report formats.

Copy picture

Sequencher can copy the images from many of its windows, including text windows, to a picture on the clipboard. Select the window you want copied, go to the **Edit** menu and choose **Copy As** then click on **Picture.**

Print setup

Before printing, go to the **File** menu and choose **Print Setup…** to change the options that are specific to your printer. Some options that may be available on your printer include portrait or landscape printing (vertical or horizontal orientation), paper size, reduction, enlargement, or special handling for color printing. (See Chapter 21, "Customizing Sequencher and User Preferences" for more information.)

Print Trace in One Page

This command appears within the **File** menu and allows you to print an individual trace so that the entire sequence fits on one page.

Printing in detail

Most windows that contain data can be printed. To print, select a window by bringing it to the front. Then go to the **File** menu and choose **Print** and enter your print specifications.

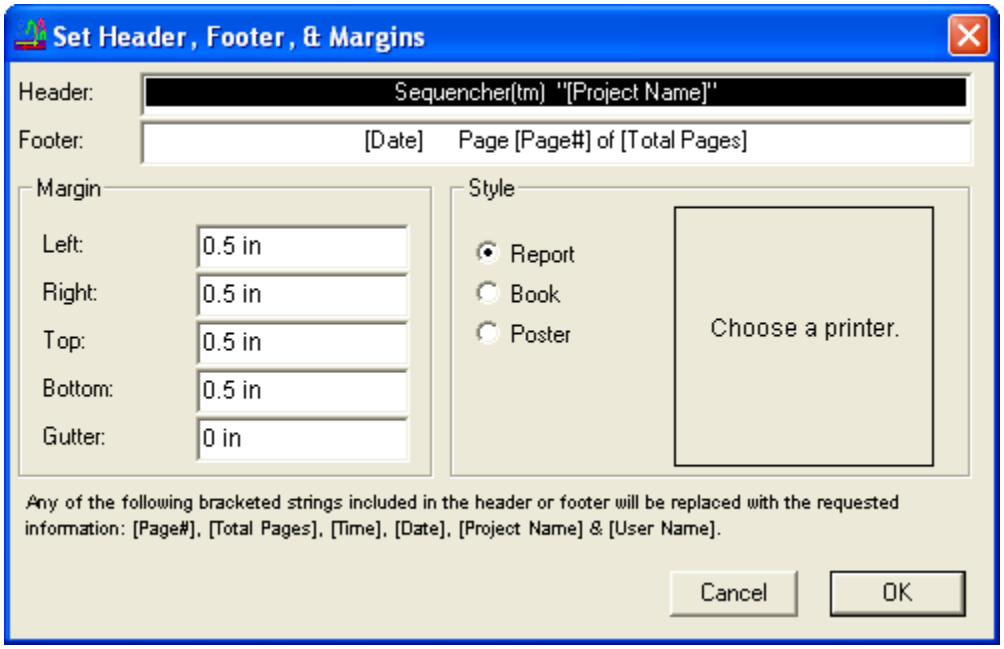
Setting header, footer and margin options

Specify header, footer and margins for any printout by going to the **File** menu and choosing **Set Header & Footer…**. You enter the text of the header and footer and a code style for elements into a dialog box (Figure 20-1) that Sequencher will automatically increase incrementally. The code elements are [Date], [Time], [Page], [Total Pages],

[Project Name] and [User Name]. You can also add your own free text in either the header or the footer. Headers and footers are supported in all reports.

Set the margins by typing appropriate numbers into the **Left, Right, Top, Bottom** and **Gutter** boxes in the **Margin** pane. Remember not to set margins smaller than those your printer can handle.

Figure 20-1 Setting margins for printouts



Report formats

Special report formats are used for printing the contents of the contig editor. Specify the report format you want by clicking the **Report, Manual** or **Poster** radio buttons from the **Style** pane of the **Set Header & Footer…** dialog box. Sequencher then adjusts the report accordingly.

The **Report** option is for reports printed on single-sided pieces of paper that need to be displayed one page at a time. Each page is fully labeled.

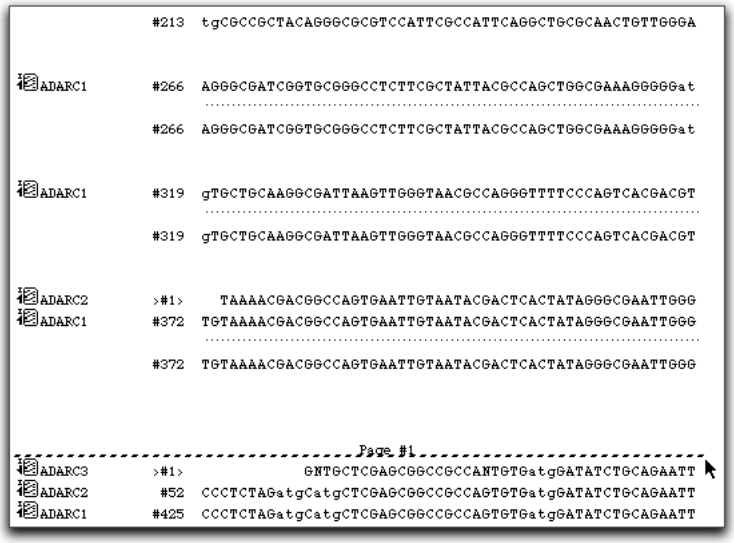
The **Manual** option is for two-sided printouts when the user can look at a left and a right page simultaneously. Only the left-hand page of each pair will have all the labels for each sequence printed in a report; the right page uses the additional space for more data.

The **Poster** option is for displaying all your data together. It minimizes the amount of labeling and assumes that the entire report is to be laid out as one big rectangle for paste-up and display.

Page breaks

Page breaks, which are shown on screen as in Figure 20-2, appear in the Sequence Editor and in the Summary Report view of a contig.

Figure 20-2 A page break in the Summary report



21. Customizing Sequencher and User Preferences

In this chapter we discuss how to set your view preferences, such as colors, custom codes, features and motifs and the genetic codes you want to use. We also discuss how to set your user preferences such as confidence scores, labels and names, assembly parameters and how you want to display chromatograms. You will learn that all of these preferences can be saved as individual or group preferences.

View Preferences

Sequencher offers you a number of options for specifying how you want to view your data in the sequence and contig editors. These options can provide visual cues which help to differentiate your data.

Ambiguous bases

Sequencher can underline all ambiguous bases in the sequence or contig editors. Go to the **View** menu and choose **Base Ambiguities As.** Go to the submenu and choose **Underlined** to mark ambiguities or **Not Highlighted** for unmarked ambiguities.

Edited Bases

Sequencher can highlight bases that have been edited, which helps you audit the integrity of your data. Go to the **View** menu and choose **Base Edits As**. Select **Not Highlighted** for unmarked edits, **Bold Magenta** for colored edits and **bOLD & cASE cHANGE** for bold and a changed case.

Display Color Bases

Sequencher can display bases in color. To turn on this feature for either a sequence or contig editor display, choose **Display Color Bases** from the **View** menu. (For information on how to change the color assignments see the section “Ambiguity Codes” in this chapter.)

Colors as Backgrounds

Sometimes, displaying bases in color does not differentiate them enough. You can further enhance the colored bases by going to the **View** menu and choosing **Colors As Backgrounds.** Sequencher then displays the base as a colored rectangle with the base character printed over it.

Display Base Confidences

Sequencher can represent imported confidence scores as a colored background. In order to view base confidence values as colored backgrounds, go to the **View** menu and click on **Display Base Confidences**. Sequencher will only display these backgrounds if the original data included base quality scores. (For information on how to change the threshold values see the section "User Preferences" in this chapter.) You may need to turn off **Colors As Backgrounds.**

Display Features

Sequencher can display features you have entered. Go to the **View** menu and click on **Display Features** to toggle this view on and off. For each individual feature you must also enable or disable its display in User Preferences. (See also Chapter 17, "Motifs and Features" for more information.)

Display Motifs

Sequencher can display motifs you have entered. Go to the **View** menu and click on **Display Motifs** to toggle this view on and off. (See Chapter 17, "Motifs and Features" for more information.)

Setting up custom codes

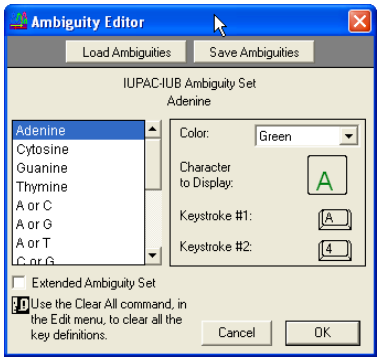
Standard coding system

The IUPAC-IUB ambiguity set is the most common coding system for representing DNA. Some commercial and academic software packages also use the Stanford Extended System codes, which include one-character representations for "maybe A," "probably G," etc. Sequencher defaults to the IUPAC-IUB system, but includes a copy of the Stanford extended set. See Appendix 25 and Appendix 0 for tables of the IUPAC-IUB and Stanford Extended code systems.

Configuring Ambiguity Codes

If you wish to configure Sequencher's ambiguity coding system yourself, go to the **Window** menu and choose **Ambiguity/Key Codes...**. Sequencher displays the dialog box shown in Figure 21-1.

Figure 21-1 Ambiguity editor



Each base and each ambiguity has a display character which is drawn on the screen and sent to the printer to represent that base or ambiguity. The display character interprets ASCII text files imported into Sequencher. Each code is also assigned a color which can be displayed in fragments or contigs by going to the **View** menu and clicking on **Display Color Bases**.

Select a base or ambiguity from the scrolling list at the left by clicking on its name. When the item is highlighted, information about it appears on the right side of the dialog box. To change the display character, click the box to highlight it.

Replacing an existing character keystroke code

Each base code can be assigned up to two alternate keys from the keyboard. Using either of these keys generates the desired base code when using the sequence or contig editors. These keystroke settings do not affect importing of data.

To change a keystroke setting for a base code, select the base code from the list on the left of the window. Click on one of the keyboard keys to highlight it and then press the key you want to assign to that base code.

To change a color setting for a base code, select the base code from the list on the left of the window. Select a different color from the **Color** drop down menu.

Saving and loading custom codes

You can load and save your customized ambiguity settings by going to the button bar at the top of the **Ambiguity Editor** and clicking on **Load Ambiguities** and then **Save Ambiguities.**

Show and hide ambiguity helper

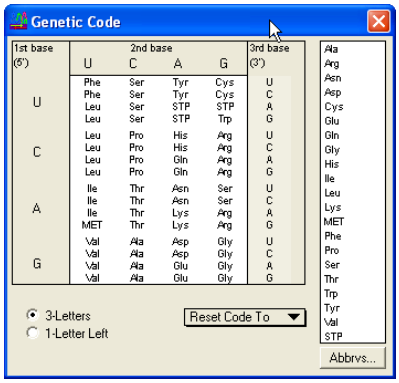
Sequencher can show a small quick-reference window that lists the current ambiguity symbols. To bring up this window, go to the **Window** menu and click on **Show Ambiguity Helper.** Close the window with the close box in the upper right corner.

Choosing a genetic code

You will find that Sequencher already has the major genetic codes built in. You can choose from among these predefined coding systems by going to the **Window** menu and choosing **Genetic Code…**. Select the code you want from **Reset Code To** pull down menu. The genetic code dialog box is shown in Figure 21-2.

Use the buttons at the bottom of the box to choose three letter or one-letter abbreviations. If you chose one-letter abbreviations, the letter will be aligned with the first (left) base of the codon.

Figure 21-2 Genetic code editor



Editing a genetic code

You can take any one of the pre-existing genetic codes and edit it in the genetic code editor to change the codon translations. The genetic code editor has two main parts: abbreviations and three-letter or one-letter codes.

Go to the **Window** menu and choose **Genetic Code**. Choose the code you want to edit and then click the table entry you want to change. Finally click the abbreviation you want in that position.

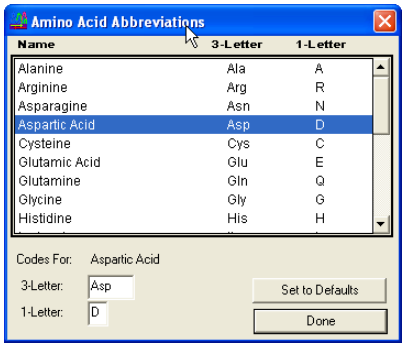
To undo any changes, select the appropriate genetic code from the **Reset Code** pull down menu.

Abbreviations

To edit the abbreviations, click the **Abbrvs** button in the bottom-right corner of the dialog box.

Another dialog box (Figure 21-3) then lets you change both the three-letter abbreviations for the amino acid names and the corresponding one-letter codes. The standard abbreviations and one-letter codes are available as the default cases.

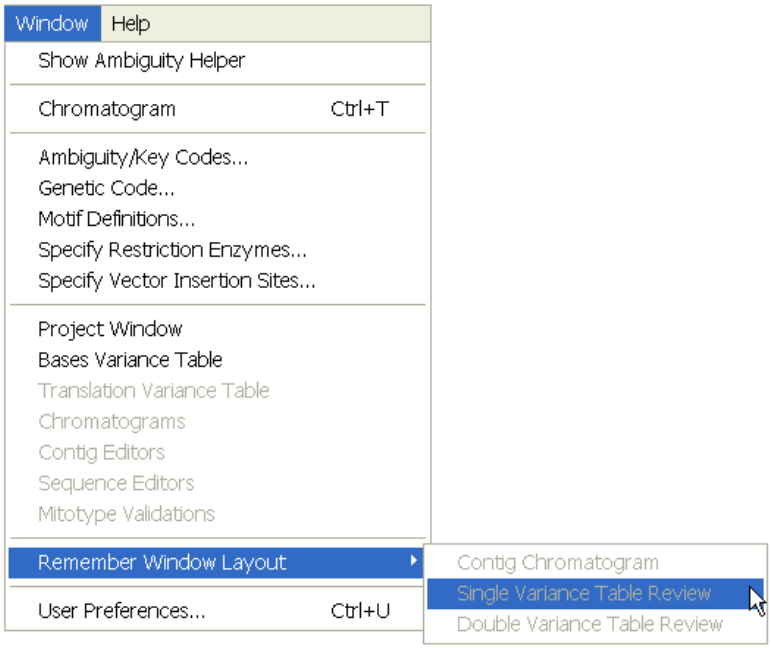
Figure 21-3 Abbreviations editor



Remember Window Layout

The **Remember Window Layout** command lets you define default positions for the Contig Editor and Contig Chromatogram when you are editing in the **Bases** view. First you must organize the windows. Then go to the **Window** menu and choose the command called **Remember Window Layout** (Figure 21-4). Select the submenu option called **Contig Chromatogram.** Every time you select the **Show Chromatograms** button Sequencher will open the windows in your layout positions.

Figure 21-4 Remember Window Layout



You can also use the **Remember Window Layout** command to define default positions for the Contig Editor and Contig Chromatogram and Variance Table or Translated Variance Table when you are in the **Review** mode. First you must organize the windows. Then go to the **Window** menu and choose the command called **Remember Window Layout.** Select the submenu option called **Single Variance Table Review** if you have one Variance Table open and **Double Variance Table Review** if both are open**.** Every

time you select the **Review** button from the Variance Table, Sequencher will open the windows in your layout positions.

User Preferences

You can specify your preferences for many program settings in Sequencher. To set these, go to the **Window** menu and choose **User Preferences….**

The **User Preferences** window contains a hierarchical list of preference settings. The main categories of settings are **General, Display** and **Input/Output**. Click the triangle sign next to a category name to expand the list and show the various preference topics within that category. When a category list is expanded the triangle changes to point downwards.

When you click a preference topic name, a Preference Pane appears on the right side of the window and enables you to specify your preferences for that topic.

Note: Each preference setting you select will be stored when you quit the program.

General

Settings

In the **File** menu you will find a command called **Open Recent**. The submenu for this command is a list of projects. You can specify the number of recent projects Sequencher will remember in this submenu. The maximum number is 99.

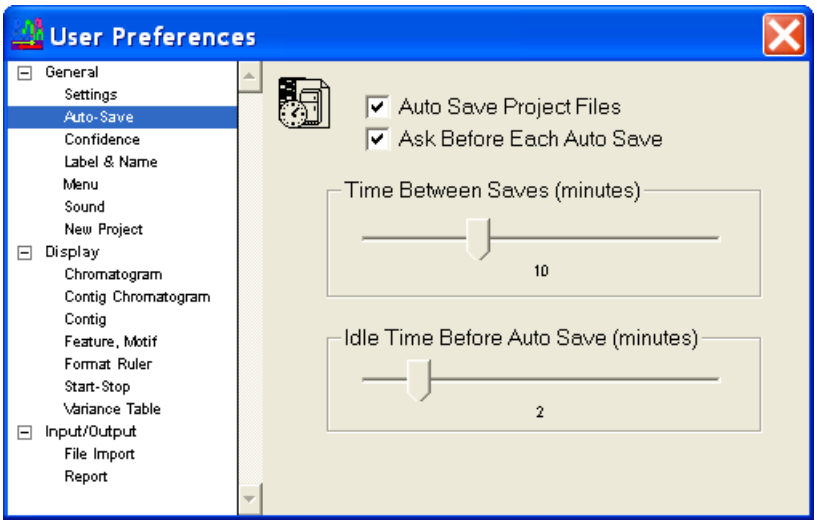
Sequencher is installed with a set of default values. As you work with the program, it replaces the default values with your preferences. However, if you want to restore the original settings, you will find a button at the bottom of the **Settings** user preferences called **Reset user preferences to factory defaults**. This option resets *all* of the user preferences to their original status.

Auto-Save

Auto-Save preferences control the timing of automatic saving of project files. To use this preference, you must have saved your data and given your project a name at least once. To activate **Auto-Save,** click the **Auto-Save Project Files** checkbox. Click the **Ask Before Each Auto-Save** checkbox if you want to confirm the save each time.

Use the **Time Between Saves** slider to set how long Sequencher should wait between each **Auto-Save.** Use the **Idle Time Before Auto-Save** slider to set how long your computer should be idle before Sequencher executes an **Auto-Save.** The latter setting will prevent Sequencher from performing an **Auto-Save** while you are typing.

Figure 21-5 Auto Save preference settings

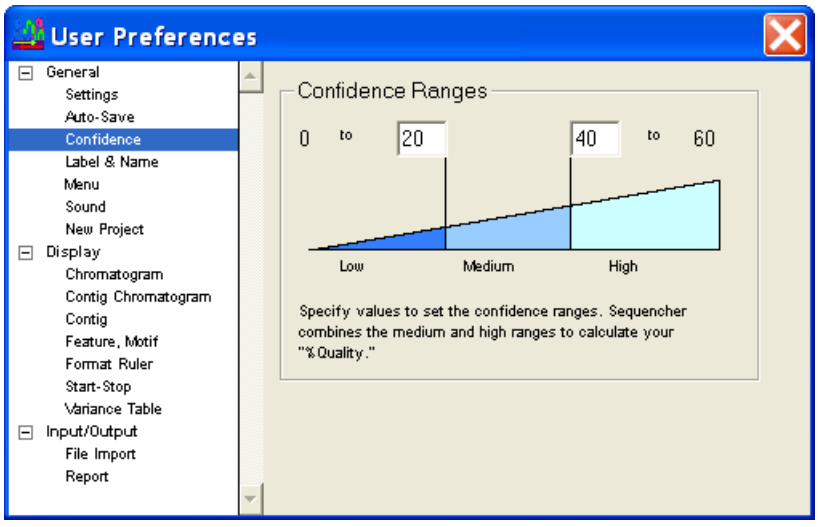


Confidence

Many of the new base callers generate confidence values associated with each base. Sequencher supports confidence values from PHRED, ACE, Trace Tuner, SCF, ABI, ESD and compatible files. The **Confidence** preferences control the display of base calls and confidence values from these files.

Sequencher's default settings show confidences below 20 as dark blue, the midrange values as a medium blue and high confidences, that is above 40, as a light blue. Change the threshold values by altering the numbers in either **Confidence Ranges** box.

Figure 21-6 Confidence Ranges settings



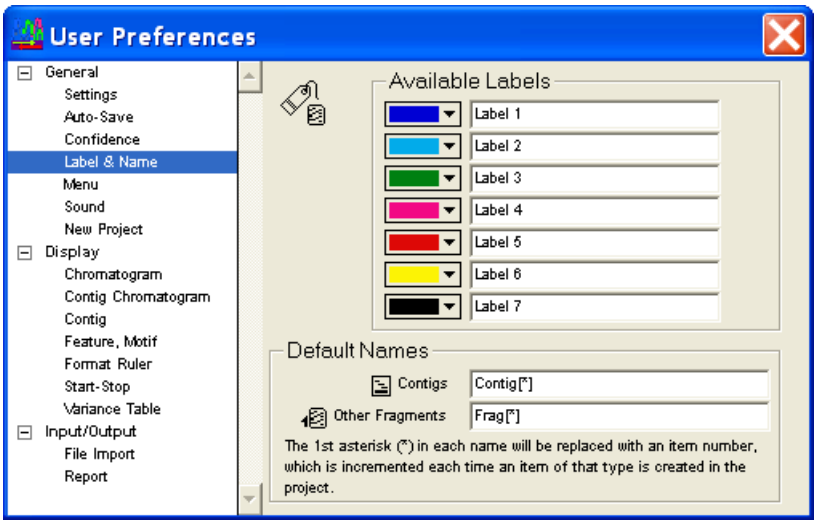
Label & Name

The **Label & Name** preferences control the labels and descriptions used to mark sequences and contigs. Use the boxes grouped under **Available Labels** to choose label colors (use the pull down color menus at the left of the label name) and descriptions for marking individual sequences.

Once you have set your preferences here you can apply them to sequences, contigs and refrigerators while you are working in a project by selecting the icon(s) in the Project Window. Go to the **Edit** menu and choose the **Label** submenu.

Under the **Default Names** section of the **Label & Name** pane, you can enter default names for the new contigs and fragments. A bracketed asterisk tells Sequencher to add a number to the name and increase it for each new contig or fragment you add to the project.

Figure 21-7 Label and name preference settings

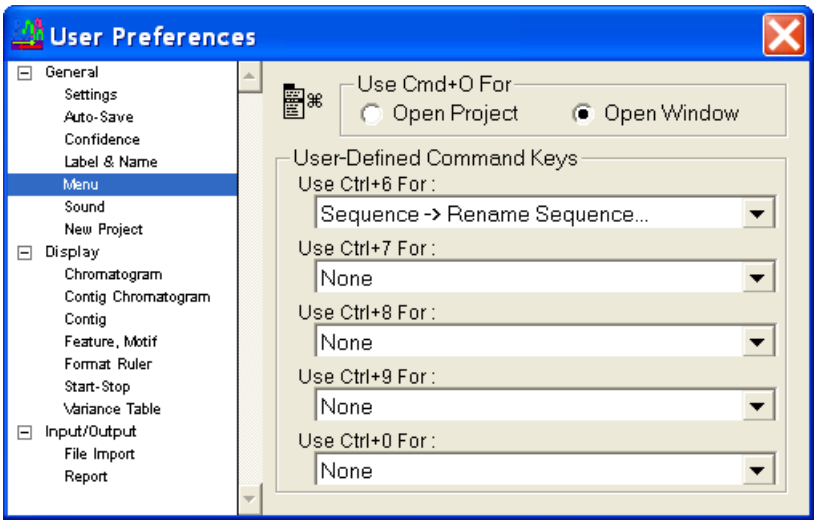


Menu

The **Menu** preferences control command options that let you optimize the way you work in your own lab. Use the buttons under **Use Ctrl+O For** to program the **Ctrl+O** key combination to open a project or open a window.

Use the pull down menus grouped under **User-Defined Command Keys** to create keyboard shortcuts for any of the menu commands. After you have assigned a keyboard shortcut to a menu item, the shortcut will be listed to the right of the menu command on the drop down menus.

Figure 21-8 Menu preference settings



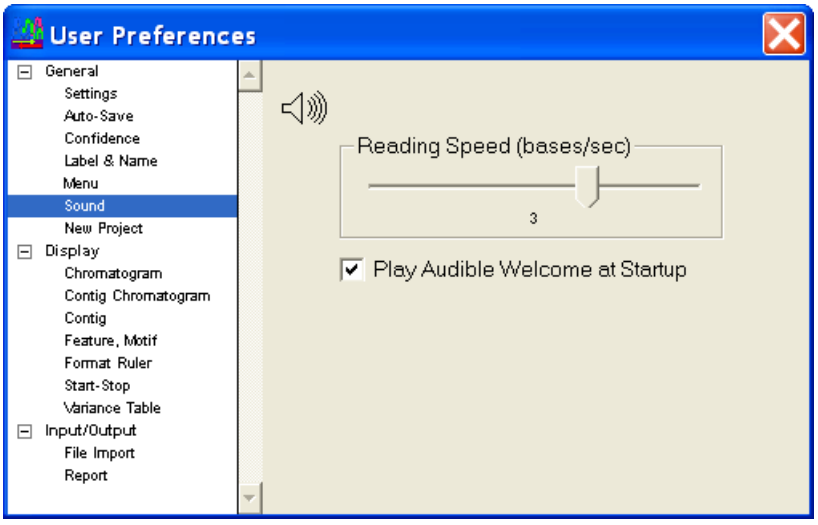
Sound

Sequencher provides audio feedback to help you maintain the quality of data input.

When you select the **Sound** icon, you will see the preference pane shown in Figure 21-9. The slider controls how quickly Sequencher reads bases to you.

Click the **Play Audible Welcome at Startup** checkbox if you want Sequencher to say "hello" at startup.

Figure 21-9 Sound preference settings



Note: You have to select a voice file to get audio output. (For more details, see Chapter 8 "The Sequence Editor" and review the section on "Voice Verification.")

New Project

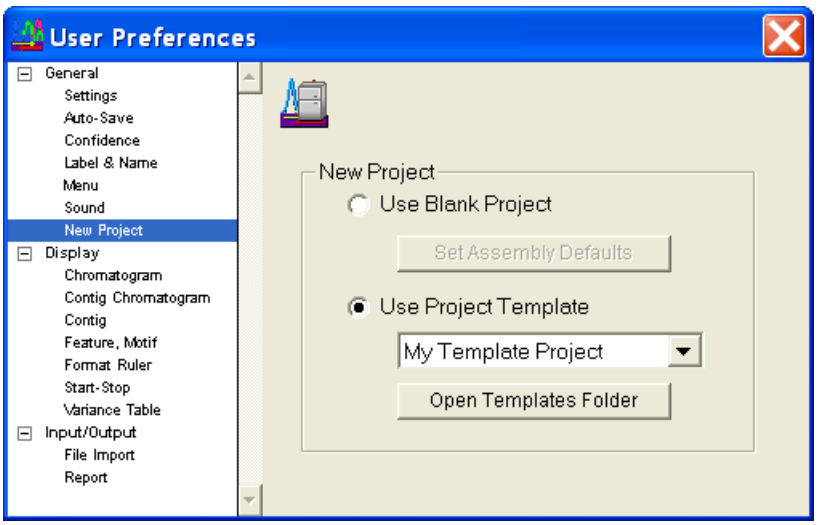
This User Preference determines what type of project will be opened by the **New Project** command.

Click the **Use Blank Project** radio button if you want your **New Project** to be blank.

Click the **Use Project Template** radio button if you want the project settings to be controlled by a template. Specify the template you want by clicking on a name in the drop down menu.

Templates are stored in a folder on your system. If you click the **Open Templates Folder** button Sequencher will open a window displaying the templates. You can store up to 1,000 templates on your system. (For more information about templates, see Chapter 3, "The Project Window.")

Figure 21-10 New Project User Preferences



Display

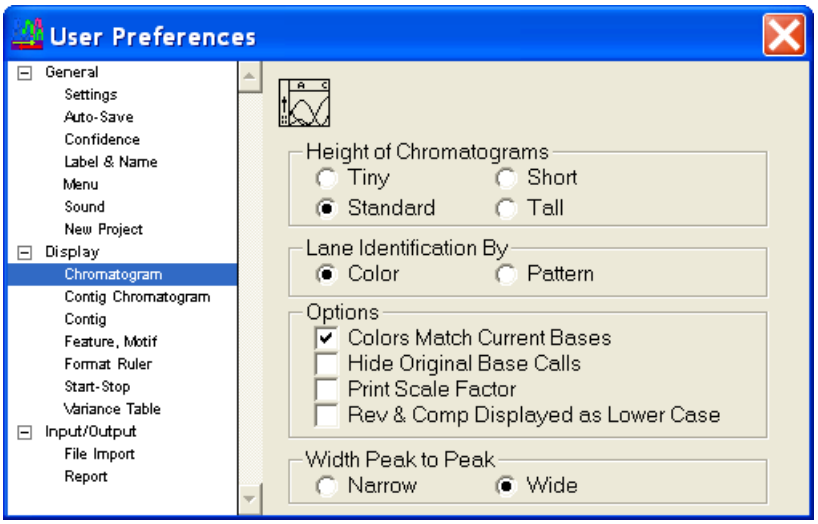
Chromatogram

Chromatogram preferences control the display of chromatograms. You can specify the **Height Of Chromatograms** by clicking one of the buttons from **Tiny** to **Tall**. If you use a black and white monitor, click the **Pattern** button for **Lane Identification By;**.

You can set lanes to be identified by color or pattern by selecting the appropriate radio button.

When viewing multiple sequences, you will notice that a reversed and complemented sequence displays the lane colors according to the original data. Click the **Colors Match Current Bases** checkbox to display all current base calls in the same color. Click the **Hide Original Base Calls** checkbox if you do not want them to appear in your chromatogram windows.

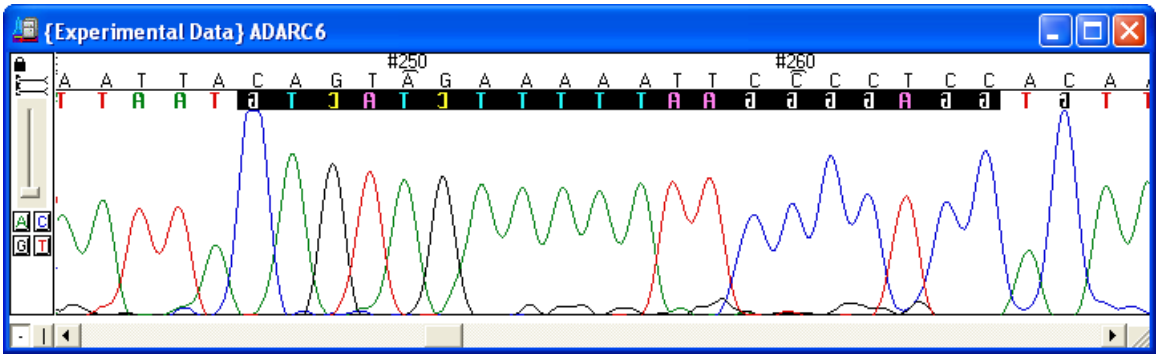
Figure 21-11 Chromatogram preference settings



If you click the **Print Scale Factor** checkbox and you used the slider shown at the left in Figure 21-12, the scale factor will appear on any printout of that chromatogram.

If you click the **Rev & Comp Displayed as Lower Case** checkbox, the original base calls for a reversed and complemented sequence are shown in lower case (as opposed to backwards) as the second of two lines of bases above the traces (as in Figure 21-12).

Figure 21-12 Reversed and complemented sequence



Under **Width Peak to Peak,** you can stretch your traces horizontally by selecting the **Wide** button.

Contig Chromatogram

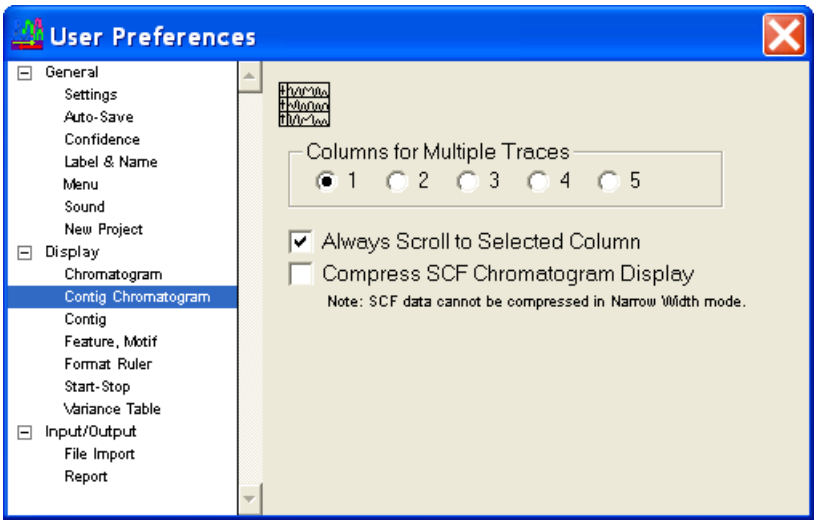
The **Contig Chromatogram** preferences control the display of chromatograms of sequences from a contig. When you have aligned several sequences from a contig and displayed the chromatograms, several traces appear in a single window, one above the other across the width of the window.

Use the **Columns For Multiple Traces** buttons to tile the data into columns so you can look at more traces at once. For example, if you select **3**, you'll see three columns of traces. If your screen and window size allow you to see three rows of traces at once, you will be able to view a total of up to nine traces without having to scroll the window.

Click the **Always Scroll To Selected Column** checkbox to have Sequencher automatically update the chromatogram window whenever you select a new column of data in the contig window.

You can have Sequencher remember your window position. Sequencher can also compress SCF chromatogram displays.

Figure 21-13 Contig chromatogram preference settings

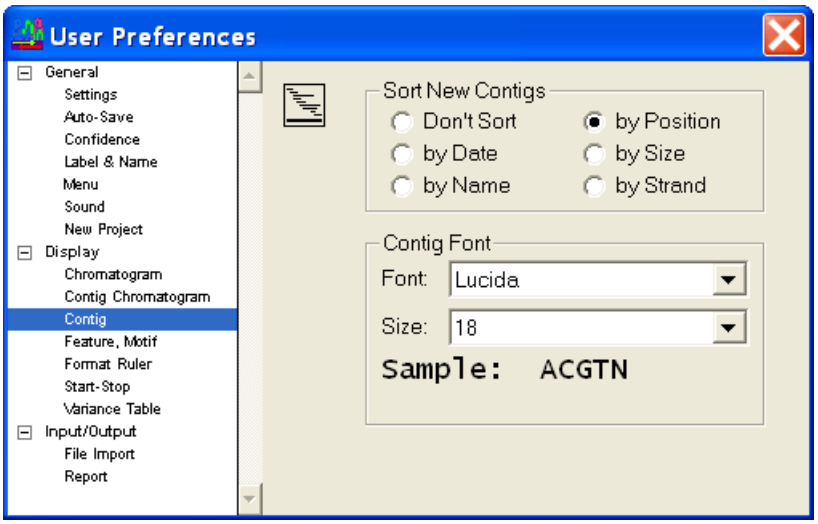


Contig

The **Contig** preferences control the vertical sorting of the fragments when a new contig is created or when fragments are added to an existing contig.

You can use the **Contig Font** option to change both the font and the font size. This allows you to use fonts other then the default Monaco 9 point font in the contig editor. You can see a sample of your current font selection below the font **Size:** drop down menu in the **Sample** portion of the preference pane.

Figure 21-14 Contig preference settings



Feature, Motif

The **Feature, Motif** preferences control the default settings for new features and motifs. When you import a sequence from GenBank into Sequencher, these settings determine the name and appearance of the feature within both the Overview and Sequencher's editor windows. (For more information on GenBank Feature tables, refer to the Appendix 28 "Feature Keys and Qualifiers".)

When you click the **Define Feature Key Default Styles…** button a new window, also called **Define Feature Key Default Styles,** appears. On the left hand side of the window is a pane listing **Feature Keys.**

To set new attributes, select the appropriate Feature Key. On the right hand side you will see a pane where you can set attributes such as the **Default** name, **Display** style and **Color**. You can restore the default settings at any time by clicking the **Restore to Defaults** button. Click the **Done** button when you have finished.

When you click on an item in the left hand **Feature Key:** pane you will see the same text appear in the **Default Name:** box. In most instances there will be extra text enclosed in square brackets. The enclosed text is the default feature qualifier. GenBank have defined several Feature Qualifiers for each Feature Key. You can replace the text within the square brackets with your preferred Feature Qualifier. (For more on Feature Qualifiers see Appendix 28 "Feature Keys and Qualifiers" and for a listing of Sequencher's default Qualifiers see Appendix 29 "Default Feature Qualifiers".)

Note: If a Label (indicated by /label= "some text") is used in a Feature Table it takes precedence over any default keys or qualifiers since it unambiguously identifies a feature or item.

Click the **Display Feature in Editors** checkbox if you want a specific feature to be displayed in the sequence or contig editors.

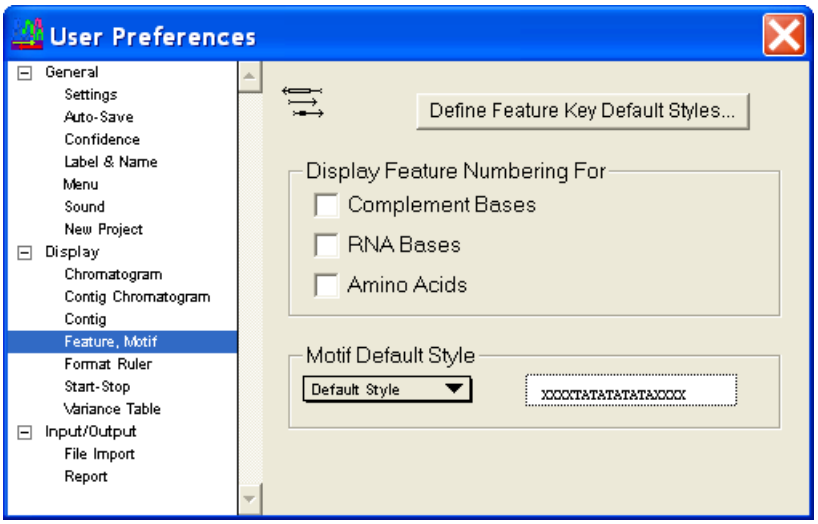
If you are making changes to the way **Feature Keys** are currently displayed you can click on the **Update Project…** button to update your project. You can decide whether to apply the current styles, name or both to **All** features in your project. If you do not want to do so, just limit your choices to **Selected** features in your project by clicking the appropriate button.

When you are finished, click on the **Done** button to return to **User Preferences.**

Use one or more of the three **Display Feature Numbering For** check boxes to indicate which of the second-line features **Complement Bases, RNA Bases**, or **Amino Acids** should have its own numbering.

Use the **Motif Default Style** pull down menu to set color, case and underlining style for new motifs. You can see a preview of the motif style in a box on the right hand side of the **Motif Default Style** preference pane.

Figure 21-15 Features and motif preference settings



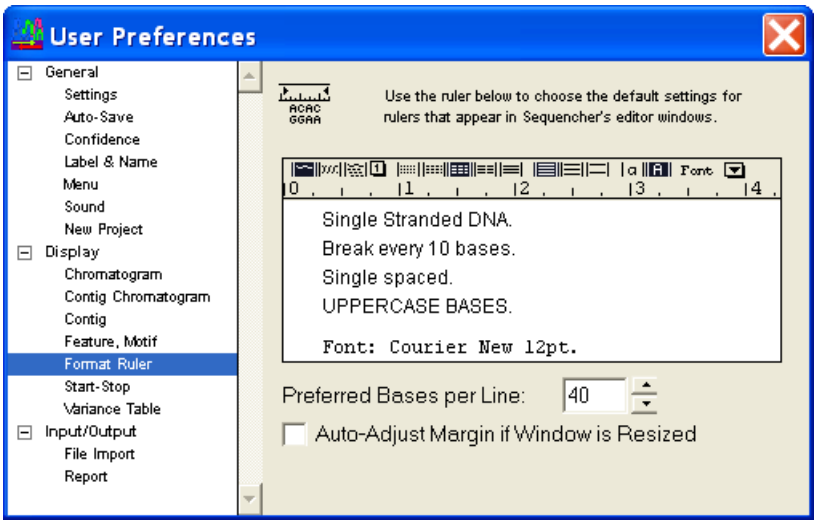
Format Ruler

The **Format Ruler** preferences controls the display defaults for any view which has a format ruler. These are just defaults; you can change these settings later from within a specific editor.

When you click an icon in the sets of icons above the ruler, your selection is explained and confirmed in the panel below the ruler (Figure 21-16). For example, use the **Preferred Bases per Line** counter to increase or decrease the preferred number of bases in each line. Note that if you request a line length that will not fit on the printed page, Sequencher uses the maximum line length that will fit.

If you want the margin to be automatically adjusted when you increase the size of a window, check the **Auto-Adjust Margin if Window is Resized** checkbox.

Figure 21-16 Format ruler preference settings

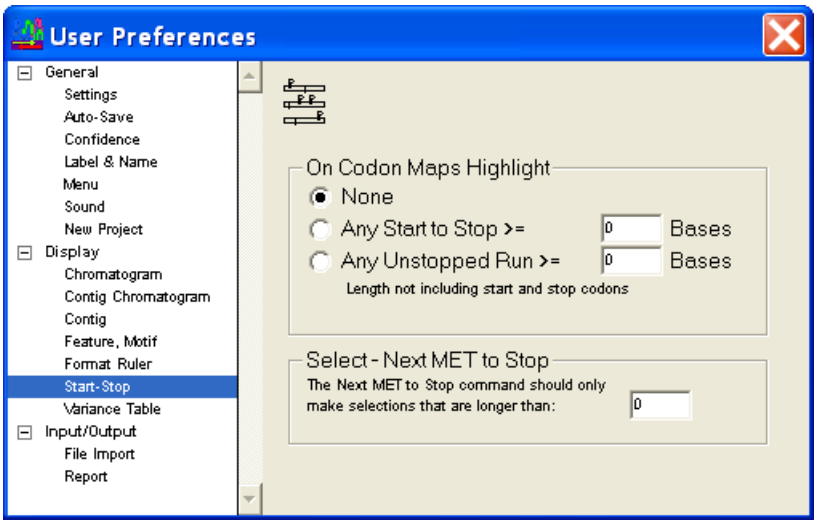


Start-Stop

The **Start-Stop** preferences control the locating of open reading frames from an editor. Use the buttons grouped under **On Codon Maps Highlight** to indicate whether you want **None,** which gives you the plain map of start and stop codons. Use **Any Start to Stop >=** or **Any Unstopped Run >=** to give you a shaded box showing open frames of a length you specify.

Figure 21-17 shows selections in the Start-Stop preference pane that will make Sequencher highlight ORFs beginning with a MET and longer than 60 bases.

Figure 21-17 Start-stop codon preference settings



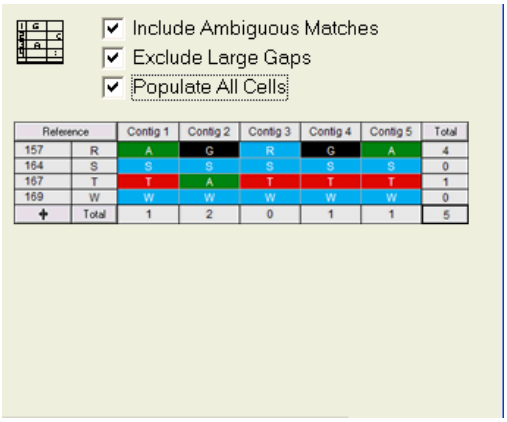
Under **Select Next MET To Stop**, you can define the minimum length of ORF selections in a Sequence Editor. If you specify a minimum length here, for example 30 bases, the **Next MET To Stop** command under the **Select** menu in the Sequence Editor will change to read **Next MET To Stop (>30b).**

Variance Table

The **Variance Table** preferences control how ambiguous matches and large gaps will be treated. If you wish to include ambiguous matches in your table, click the **Include Ambiguous Matches** checkbox. If you wish to exclude large gaps from your table, click the **Exclude Large Gaps** checkbox.

When you generate a Variance Table you would expect to have blank cells where the base at that position matches the exemplar. If you would like to see all bases across the comparison range whether they match or are different to the exemplar then click the **Populate All Cells** checkbox (see Figure 21-18).

Figure 21-18 Variance Table preference settings


Include Ambiguous Matches
Exclude Large Gaps
Populate All Cells
Reference
Contig 1
Contig 2
Contig 3
Contig 4
Contig 5
Total
157 R A G R G A 4
164 S S S S S S 0
167 T T A T T T 1
169 W W W W W W 0
Total 1 2 0 1 1 5

Input/Output

File Import

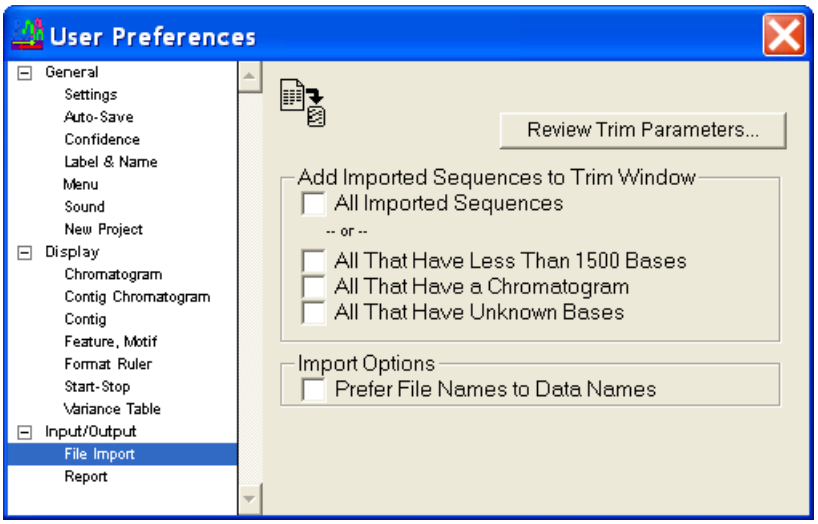
File Import preferences control how Sequencher applies the automatic trim function to imported files.

Use the checkboxes grouped under **Add Imported Sequences To Trim Window** to specify which sequences Sequencher should examine for poor quality ends and then trim.

You can also specify how you want to resolve conflicts between the names of imported files and names stored with particular sequences. If you click the **Prefer File Names To Data Names** checkbox, Sequencher uses the file name.

Clicking the **Review Trim Parameters** button displays the **Ends Trimming** dialog box, in which you can set trim criteria. (See Chapter 6, "Preparing your Data for Assembly," for details.)

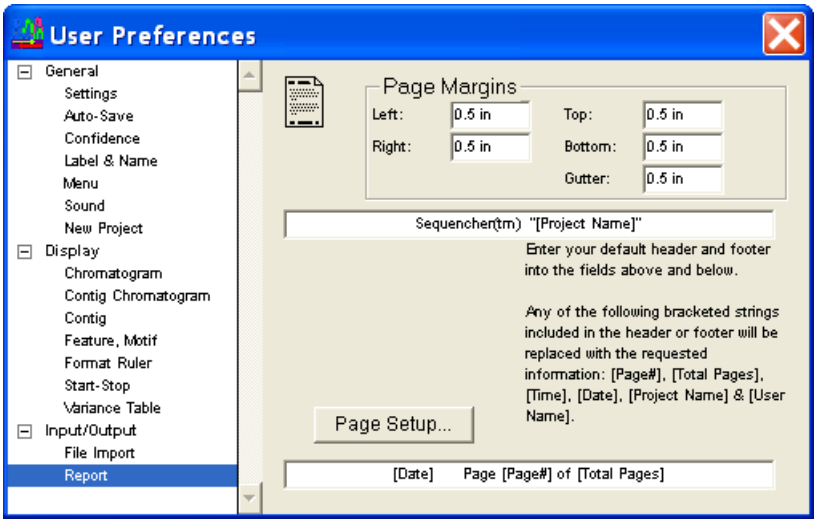
Figure 21-19 File import preference settings



Report

The **Report** preferences command controls the default page setup for report printouts. Use the fields in the **Page Margins** box to set the margins and gutter. Click **Page Setup** to open the printer page setup dialog box. The page setup settings will apply to all windows you open thereafter.

Figure 21-20 Report margin preference settings



You can specify headers and footers on reports generated from any window in which you are working. Any text you type into the field above the page layout will appear in the header and text you type in the field below the layout will appear in the footer.

For variable elements (date, time, page and total number of pages), enclose the word—typed with an initial capital—in square brackets, as shown in Figure 21-20 above.

22. Forensic Features

If you have the forensic version of Sequencher, you have access to special functions for DNA based identification. In this chapter we describe how you can validate mtDNA, create reports and export your results.

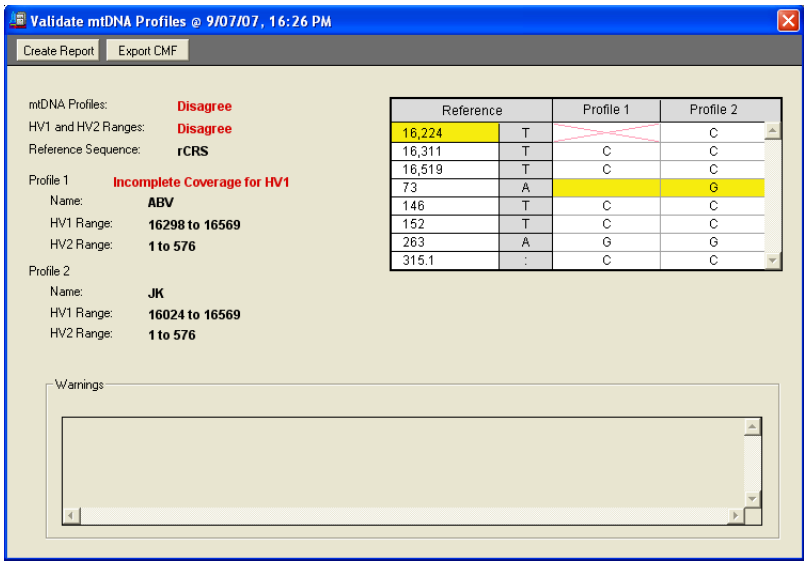
Working with mtDNA profiles

Validating mitochondrial DNA profiles

The **Validate mtDNA Profiles** command allows you to compare the results of separate analysts. Select two contigs which have been assembled to the same Reference Sequence. Go to the **Contig** menu and click on the **Validate mtDNA Profiles** item. This command creates a report in a new window. The title bar includes a date and time stamp as shown in Figure 22-1 below. The report includes the name of the Reference Sequence used, the name of the contigs being examined and the HV1 and HV2 ranges. Underneath the main body of the report is a text box that automatically displays pertinent warnings.

In the report shown in Figure 22-1, the two mtDNA profiles disagree. The bases in conflict are highlighted in yellow. Also there are positions where there is data for one mtDNA profile, but not for the other. The base position is distinguished by a yellow highlight, and the cell with the missing data has a pink X though it. The cells are unshaded where the two profiles both agree and confirm each other.

Figure 22-1 The Validate mtDNA Profiles window



Create Reports

At the top of the Validate mtDNA window is a button bar. Click on the **Create Report** button. A new window called Analyst Report opens. This contains a version of the Validate mtDNA report that you can send to a printer by clicking the **Print** button. You may also save this report by clicking on the **Save** button.

Note: The report includes spaces for the analysts' and reviewer's signatures.

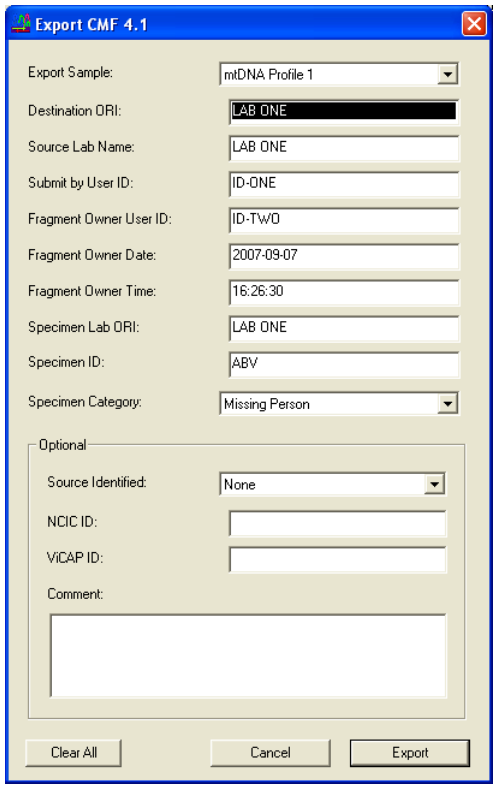
Export CMF

You can create a CMF report by clicking on the **Export CMF** button. A new window called Export CMF will open. Choose the **Export Sample:** and **Specimen Category:** from the relevant drop down menus. The **Fragment Owner Date:**, the **Fragment Owner, Time:** and **Specimen ID:** fields are filled in automatically. You will need to enter information into the remaining text fields.

There are a number of optional fields. The **Source Identified:** field is a drop down menu. There are two text fields for the **NCIC ID** and **ViCAP ID**. There is also a free text **Comment:** field. When you are finished click on the **Export** button. If you have made an error you can click on the **Clear All** button to reset the form or you can click the **Cancel** button to return to the Validate MtDNA window.

Note: If you move your cursor to pause over a field, you will see a brief description summarizing the information required for that field.

Figure 22-2 Export CMF report



Further Commands of Interest

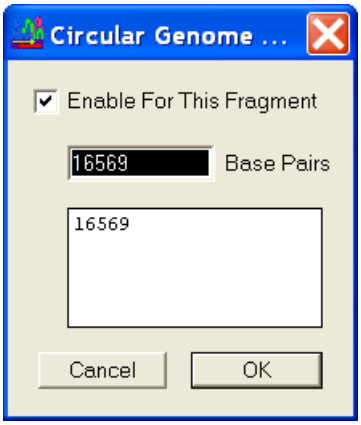
Set Circular Genome Size

The **Set Circular Genome Size…** command allows you to set the number of bases in your DNA circle.

Select a sequence that has already been defined as a Reference Sequence. Then go to the **Sequence** menu and choose the **Set Circular Genome Size…** command. The dialog shown in Figure 22-3 below will appear. Enable the circular number by clicking in the checkbox. If you have enabled the Cambridge Reference Sequence as your Reference Sequence, the number of base pairs in your fragment will automatically be set to 16,569. Dismiss the dialog window by clicking the **OK** button.

Note: Circular numbering can only be used in conjunction with a Reference Sequence. You must designate your sequence as a Reference Sequence before you can enable circular numbering.

Figure 22-3 The Circular Genome Size dialog window



How to mark a sequence as a Reference Sequence

To designate a sequence as a Reference Sequence, select the sequence icon in the Project Window and under the **Sequence** menu select the **Reference Sequence** command**.** From now on, that sequence icon will include a small letter "R" to remind you that it has been marked as a Reference.

For more information on Reference Sequence properties and how to work with a Reference Sequence, see Chapter 7 “The Reference Sequence” for more information.

New Project From Template

The **New Project From Template** command allows you to open a new project containing all the sequences, settings and preferences associated with your chosen template. Go to the **File** menu and click on **New Project From Template**. Select a template from the submenu. A new blank Project Window will open.

Note: A template called rCRS is provided. This contains the Revised Cambridge Reference Sequence and is the default template for this command.

For more information on Templates see Chapter 3 “The Project Window.” There is a tutorial on analyzing mtDNA with Sequencher in the Tutorials folder.

Display secondary peak

You can identify heterozygotes by identifying the second-highest peak beneath the primary peak. If you have a particular candidate for a heterozygote, first open the chromatogram for the individual sequence then click on the base call(s).

Go to the **Sequence** menu and choose **Call Secondary Peaks…**. A dialog box lets you specify how the secondary peak should be called.

The slider lets you specify how significant the second-highest peak must be to generate a change.

If you want Ns to be replaced then select the **Allow Ns to be replaced** checkbox.

If you've already edited bases by hand and don't want them changed, you must *deselect* the checkbox called **Allow edited bases to be replaced**.

If you select **Only make changes that result in an ambiguity,** Sequencher may change an ambiguous base call to an A, C, G or T if it does not meet the secondary peak criteria. If you prefer it to remain ambiguous do not select this option.

If you wish to search just a range of bases, highlight them and check the **Search Selection Only** checkbox.

For more information see Chapter 15"Chromatograms".

23. Appendix - Keyboard shortcuts

Menu/Window	Command	Windows
Any view	Help	F1
File menu	Open window or Open project (set in User Preferences)	Ctrl+ O
	Close window	Ctrl+ W
	Get Info…	Ctrl+ I
	Save Project	Ctrl+ S
	Print	Ctrl+ P
	Exit	Ctrl+ Q
Edit menu	Undo	Ctrl+ Z
	Cut Selection	Ctrl+ X
	Copy Selection	Ctrl+ C
	Paste	Ctrl+ V
	Edit Comments…	Ctrl+,
Select menu	Select All	Ctrl+ A
	Select None	Ctrl+ \
	Find bases…	Ctrl+ F
	Find bases again	Ctrl+ G
	Contig column	Ctrl+ K
	Select Next Ambiguous base	Ctrl+ N
	Select Next Contig Disagree	Ctrl+ D
	Select Next Edited Base	Ctrl+ E

Menu/Window	Command	Windows
	Select Next Low Confidence Base	Ctrl+ L
	Select Next Met to Stop	Ctrl+ M
Sequence menu	Mark Selection As Feature	Ctrl+ ‘
View menu	Display Format Ruler	Ctrl+ R
	Reverse & Comp	Ctrl+4
	Sequenced Strand	Ctrl+5
View submenu	Display 1st frame translation	Ctrl+ 1
	Display 2nd frame translation	Ctrl+ 2
	Display 3rd frame translation	Ctrl+ 3
***	Single stranded	Ctrl+ -
	Double stranded	Ctrl+ =
Window menu	Chromatogram	Ctrl+ T
	User Preferences	Ctrl+ U
	Switch between large icons, small icons and list view	Alt click in Title line
Overview Options	Names at Left	Ctrl+ Shift <-
Sequence editor/Contig editor	Insertion point moves one base to the right	Right arrow key
	Insertion point moves one base to the left	Left arrow key
	Insertion point moves 3 bases to the right	Alt right arrow key
	Insertion point moves 3 bases to the left	Alt left arrow key
	Insertion point moves to 3’ end of sequence	Ctrl+ right key
	Insertion point moves to 5’ end of sequence	Ctrl+ left arrow key
Summary view	Display Formatting Ruler	Ctrl+ Alt R
	Frame 1	Ctrl+ Alt 1
	Frame 2	Ctrl+ Alt 2
	Frame 3	Ctrl+ Alt 3
User defined		Ctrl+6
		Ctrl+7
		Ctrl+8
		Ctrl+9
		Ctrl+0

24. Appendix – Advanced Exprssions

A regular expression is a special way of describing a search pattern using text strings according to certain syntax rules. Sequencher uses regular expressions to help you break down your sequence naming schema in order to define your Assembly Handles.

Table of Regular expression special characters

.	Represents any character.
[A-Za-z]	Represents any letter.
[A-Z].	Represents any capital letter
[a-z].	Represents any lowercase letter
\d	Represents any digit
[0-9]	Represents any digit
*	A character or special character which is followed by * matches *zero* or *more* instances of the character
+	A character or special character which is followed by **+** matches *one* or *more* instances of the character
?	A character or special character which is followed by **?** matches *zero* or *one* instances of the character
{N}	A character or special character followed by **{**N**}**, where "N" = a whole number, matches that number of instances of the preceding character
[ ]	The contents of square brackets describe a list of matching items, regardless of order.
^	Use this in square brackets to *exclude* the other items contained within the brackets
\	The \ is known as the escape character and it allows the character to be interpreted literally. It can also give a standard character a special meaning as in **\d**
( )	Matches whatever you type within the parentheses.
|	The | acts as an "or"

You may require a more complex delimiter then those provided in the **Name Delimiters** drop down menu. You can use regular expressions to define these more complex delimiters.

Table 24-1 Examples of delimiter expressions

#	Delimiter Expression	Sequence Name	Handle1	Handle2	Handle3	Handle4	Handle5	Handle6
1	%	Gel?%A%T7?%dog	Gel?	A	T7?	dog		
2	\?	Gel?%A%T7?%dog	Gel	%A%T7	%dog			
3	\?%	Gel?%A%T7?%dog	Gel	A%T7	dog			
4	\?|%	Gel?%A%T7?%dog	Gel		A	T7		dog

If your sequence name is complex you may need to write a regular expression that describes the delimiters and the Assembly Handles. In the example below the text between the parentheses defines the Assembly Handles. Anything not enclosed by parentheses will be treated as a delimiter.

(Assembly Handle 1**)**Delimiter**(**Assembly Handle 2**)**Delimiter**(**Assembly Handle 3**)**

You would use this kind of expression with the **Expression is a delimiter** box unchecked.

Regular expressions are powerful tools. There are many resources on the internet which describe how to use regular expressions. Here is one to help you get started.

http://www.macadamian.com/column/regular_expressions.html

25. Appendix - IUPAC-IUB Standard Code

Code	Base(s)	Meaning
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
U	U	Uracil (RNA)
R	A or G	Purine
Y	C or T (or U)	Pyrimidine
K	G or T (or U)	Keto
M	A or C	Amino
S	G or C	Strong
W	A or T (or U)	Weak
B	C, G, T (or U)	not A
D	G, A, T (or U)	not C
H	A, C, T (or U)	not G
V	A, C, or G	not T
N	A, C, G, T (or U)	(Any)

26. Appendix - Stanford Extended Code

Code	Base(s)	Meaning
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
U	U	Uracil (RNA)
P	A or G	Purine
Q	C or T (or U)	Pyrimidine
K	G or T (or U)	Keto
J	A or C	
M	G or C	
L	A or T (or U)	
B	C, G, T (or U)	not A
D	G, A, T (or U)	not C
H	A, C, T (or U)	not G
V	A, C, or G	not T

Code	Base(s)	Meaning
5	A or Another Base	Probably A
3	C or Another Base	Probably C
6	G or Another Base	Probably G
4	T or Another Base	Probably T
9	A or Nothing	Maybe A
7	C or Nothing	Maybe C
0	G or Nothing	Maybe G
8	T or Nothing	Maybe T
N	A, C, G, T (or U)	(Any)

27. Appendix – Default Feature Styles

The following table displays the defaults settings for the colors and styles of new features. New features are those you create in an existing sequence or those described in an imported sequences with a GenBank style Feature Table.

You can change the default settings in the Feature, Motif preference pane. (For more information, refer to Chapter 21 " Customizing Sequencher and User Preferences".

Feature	Display	Color	Style	Second Strand
A. misc_feature				
1. misc_difference a) conflict b) unsure c) old_sequence d) variation e) modified_base f) mutation g) allele	√ √ √ √ √ √ √ √	Red Red Red Red Red Red Red Red	Underlined	
2. gene				
3. misc_signal a) promoter 1) CAAT_signal 2) TATA_signal 3) -35_signal 4) -10_signal 5) GC_signal b) RBS c) polyA_signal d) enhancer e) attenuator				

Feature	Display	Color	Style	Second Strand
f) terminator g) rep_origin h) oriT				
4. misc_RNA a) prim_transcript 1) precursor_RNA a) mRNA b) 5'clip c) 3'clip d) 5'UTR e) 3'UTR f) exon g) CDS 1) sig_peptide 2) transit_peptide 3) mat_peptide h) intron i) polyA_site j) rRNA k) tRNA l) scRNA m) snRNA n) snoRNA	 √ √ √	 Blue Blue Cyan	 Invert case	 Protein
5. Immunogobulin related a) C_region b) D_segment c) J_segment d) N_region e) S_region f) V_region g) V_segment				
6. repeat_region a) repeat_unit b) LTR c) satellite d) transposon				
7. misc_binding a) primer_bind b) protein_bind c) STS d) primer				
8. misc_recomb a) iDNA				

Feature	Display	Color	Style	Second Strand
9. misc_structure a) stem_loop b) D-loop				
10. gap				
11. operon				
12. source				
-				
Unrecognized*				

**Note* – The Unrecognized key is at the same level at misc_feature.

28. Appendix - Feature Keys and Qualifiers

GenBank Feature tables

GenBank Feature Tables provide a way of describing and locating features in a DNA sequence using internationally agreed-upon layout and vocabulary. Although these tables look complex, all tables are comprised of three basic elements which are described in the table below.

Element	Description
Feature Key	A single word or abbreviation which indicates a functional grouping
Location	The instruction for where to find the feature in the sequence
Qualifier	Additional information about the feature

Feature tables may look simple or complex depending on the number of annotations in the table. The following figure shows a typical entry.

Figure 28-1 A simple feature table entry

```
Key             Location/Qualifiers
source          1..1509
                /organism="Mus musculus"
                /strain="CD1"
```

In the example above, the feature key is “source” and the location is represented by a range “1…1509”. A location can be a single base, a range of bases, a single base within a range etc. The qualifier, which can be free text, controlled vocabulary, citation or

references numbers, sequences, or user-defined feature labels is separated from the location by a forward slash.

For a more detailed discussion of table entries and their standard typography, see http://www.ncbi.nlm.nih.gov/collab/FT/index.html.

The table below contains a listing of the standard feature keys which have been implemented in Sequencher.

Table 28-1 Feature keys groups and hierarchies

Feature	Specificity or detail	Functional grouping
A. misc_feature		
1. misc_difference	a) conflict b) unsure c) old_sequence d) variation e) modified_base f) mutation g) allele	Difference and change features
2. gene		
3. misc_signal	a) promoter 1) CAAT_signal 2) TATA_signal 3) -35_signal 4) -10_signal 5) GC_signal b) RBS c) polyA_signal d) enhancer e) attenuator f) terminator g) rep_origin h) oriT	Expression signal features
4. misc_RNA	a) prim_transcript 1) precursor_RNA a) mRNA b) 5'clip c) 3'clip d) 5'UTR e) 3'UTR f) exon g) CDS 1) sig_peptide 2) transit_peptide 3) mat_peptide h) intron	Transcript features

Feature	Specificity or detail	Functional grouping
	i) polyA_site j) rRNA k) tRNA l) scRNA m) snRNA n) snoRNA	
5. Immunogobulin related	a) C_region b) D_segment c) J_segment d) N_region e) S_region f) V_region g) V_segment	
6. repeat_region	a) repeat_unit b) LTR c) satellite d) transposon	Repeat features
7. misc_binding	a) primer_bind b) protein_bind c) STS d) primer	Binding features
8. misc_recomb	a) iDNA	Recombination features
9. misc_structure	a) stem_loop b) D-loop	Structure features
10. gap		
11. operon		
12. source		

Feature table qualifiers

Genbank Feature Qualifiers convey many different types of information. To accommodate this breadth there are several value formats:

Table 28-2 Feature Qualifiers

Qualifier Type	Description
Free text	Free text qualifiers are usually descriptive phrases enclosed in double quotation marks.
Controlled vocabulary or enumerated values	Controlled vocabulary or enumerated values qualifiers are selections from a controlled vocabulary list and are entered without quotation marks. Examples of controlled vocabularies can be found in the Appendices to the GenBank Feature Table documentation (http://www.ncbi.nlm.nih.gov/collab/FT/index.html).
Citation or reference numbers	A citation or reference number is enclosed in square

Qualifier Type	Description
	brackets to distinguish it from other numbers.
Sequences	Since 1998, it has not been acceptable to use a literal sequence of bases as the Sequence Qualifier.
Feature labels	The feature label qualifier supports clarity in reporting by making sure your references are unambiguous. An example of a feature label would be alcA (see example below).

The figure below shows a typical example of a feature table entry with qualifiers. In this example the feature key is CDS. The location is expressed as two numbers. The qualifiers tell the reader that the gene is alcA, the product of the gene is alcohol dehydrogenase and that the evidence is experimental.

Figure 28-2 Qualifier examples

```
Key             Location/Qualifiers

CDS             6006..>6253
                /gene="alcA"
                /experiment="experimental evidence, no additional details
                recorded"
                /note="ADHI; ethanol regulon"
                /codon_start=1
                /product="alcohol dehydrogenase I"
```

Information on the Feature Keys and Tables and Qualifiers has been taken from the official GenBank documentation which can be obtained from the following URL http://www.ncbi.nlm.nih.gov/collab/FT/index.html.

29. Appendix - Default Feature Qualifiers

In this table you will find the default Feature Qualifier for each Feature Key as implemented in Sequencher.

Feature Key	Default Feature Qualifier
Sequencher *	
A. misc_feature	misc feature [note]
1. misc_difference	misc difference [note]
a) conflict	conflict [compare]
b) unsure	unsure difference
c) old_sequence	old sequence [compare]
d) variation	variation [note]
e) modified_base	[gene] modified base [mod_base]
Allele **	allele
Mutation **	mutation
2. gene	[gene] gene
3. misc_signal	[gene] misc signal
a) promoter	[gene] promoter
1) CAAT_signal	[gene] CAAT signal
2) TATA_signal	[gene] TATA signal
3) -35_signal	[gene] -35 signal
4) -10_signal	[gene] -10 signal
5) GC_signal	[gene] GC signal
b) RBS	[gene] RBS
c) polyA_signal	[gene] polyA signal
d) enhancer	[gene] enhancer
e) attenuator	[gene] attenuator

Feature Key	Default Feature Qualifier
f) terminator	[gene] terminator
g) rep_origin	[gene] replication origin
h) oriT	[gene] oriT
4. misc_RNA	[gene] misc RNA
a) prim_transcript	[gene] primary transcript
1) precursor_RNA	[gene] precursor RNA
a) mRNA	[gene] mRNA
b) 5'clip	[gene] 5'clip
c) 3'clip	[gene] 3'clip
d) 5'UTR	[gene] 5'UTR
e) 3'UTR	[gene] 3'UTR
f) exon	[gene] exon
g) CDS	[gene] CDS
1) sig_peptide	[gene] signal peptide
2) transit_peptide	[gene] transit peptide
3) mat_peptide	[gene] mature peptide
h) intron	[gene] intron
i) polyA_site	[gene] polyA site
j) rRNA	[gene] rRNA
k) tRNA	[gene] tRNA
l) scRNA	[gene] scRNA
m) snRNA	[gene] snRNA
n) snoRNA	[gene] snoRNA
5. Immunogobulin related	Immunogobulin related
a) C_region	[gene] C region
b) D_segment	[gene] D segment
c) J_segment	[gene] J segment
d) N_region	[gene] N region
e) S_region	[gene] S region
f) V_region	[gene] V region
g) V_segment	[gene] V segment
6. repeat_region	[gene] [rpt_type] repeat region [rpt_family]
a) repeat_unit	[gene] [rpt_type] repeat unit
b) LTR	[gene] LTR
c) satellite	[gene] [rpt_type] satellite
Transposon **	transposon
7. misc_binding	misc binding of [bound_moiety]
a) primer_bind	[gene] primer binding site
b) STS	STS [locus_tag]
c) protein_bind	[bound_moiety] protein binding site
Primer**	primer
8. misc_recomb	Misc recombination [gene] [map]
a) iDNA	[gene] iDNA
9. misc_structure	misc structure [note]

Feature Key	Default Feature Qualifier
a) stem_loop	[gene] stem loop
b) D-loop	[gene] D-loop
10. gap	gap of [estimated_length]
11. operon	[operon] operon
Source	[organism]
Unrecognized ***	

* Sequencher Feature Key is for personal annotation

** Legacy keys

*** This is used for Feature Keys which are unknown to Sequencher so that they can be imported in to a project

30. Glossary

Name	Definition
Administrator	Also known as Admin or admin user. A user account that typically has privileges which allow the user to change system settings and install software.
Agent	A region in some dialog boxes (and also in the contig editor) that contains information in text form.
ASCII	American Standard Code for Information Interchange; the most widely accepted coding system that permits letters, numbers and punctuation to be represented in binary computer code.
Assemble by Name	A feature which uses portions of sequences's name to determine against which other sequences it will be compared against to construct a contig.
Base caller	Program that analyses fluorescent trace intensities from automated sequencer and assigns bases together with a probability of error.
Button bar	The tool region in a Sequencher window. The button bar is located just below the title bar.
CAF	A format for describing sequence assemblies.
Checkbox	A control that lets you toggle a parameter. Click the box to turn it on or off. When a checkbox is active, it is displayed with an "X" in the box.
Click	Press and immediately release the mouse button.
Confidence score	A number associated with each base call and which defines the likelihood that a base call is incorrect. The most common scale is from 1-60, where "60" represents a $1/10^6$ chance of a wrong call and 20 represents a $1/10^2$ chance. See also Quality score.

Name	Definition
Contig	Contiguous alignment of a set of sequences to make up the sequence of bases from a longer piece of DNA.
Cursor	The indicator on a computer screen that shows the currently active location. Cursors have various shapes, indicating different purposes and capabilities.
Dialog box	Any of several special windows that the operating system or Sequencher displays when it needs to alert you to something you need to know, or needs to ask you a question.
Delimiter	Usually a character such as / or – which separates elements of a sequence name.
Double-click	Click twice in quick succession. You can set the amount of time you have before the machine decides that you have performed a second single click rather than a double-click.
Drag	Press the mouse button and hold it down while you move the mouse (and thus the cursor) across the screen.
Elevator button	A very small scroller that allows you to increase or decrease a numerical value by using the mouse. An elevator button has an up arrow and a down arrow.
Enzyme	In this manual, enzyme always refers to a restriction endonuclease.
Exemplar	Chosen or primary sequence. Literal meaning typical example or excellent model.
Field	A rectangle in which the user can (or in some cases must) enter names or other information.
Handle	The portion of a sequence name between two consecutive delimiters. Also known as Assembly handle.
Hierarchical menu	A hidden menu that looks like a command until you invoke it. When you move the cursor over the name of the hierarchical menu, it appears to the right of the parent menu. A hierarchical menu is indicated by a right-pointing triangle.
Modal dialog box	A dialog box that requires a response before you can continue with regular actions in the application.
Motif	In Sequencher this is a short subsequence of 50 bases or less.
Mouse	The mouse is integral to the interface. It controls the position of the cursor on the screen and is used to enter some kinds of information. (See Click, Double-click, Drag).
MSF	Multiple Sequence Format. A format used by multiple sequence

Name	Definition
	alignment programs.
NEXUS/Paup	Format used by phylogentic and cladistic software.
Project	A project contains sequences, information on how these sequences fit together to form larger pieces and parameter information, specified by the user, that controls alignment operations.
Project Window	The window in which Sequencher displays the contigs and unincorporated Sequences of a project. Items in the Project Window are can be shown as icons, small icons or as a sorted list.
Quality score	A number associated with each base call and which defines the likelihood that a base call is incorrect. See also Confidence score.
Radio button	A control that lets you choose among alternate possibilities. Radio buttons always appear in sets of at least two. Click one of the buttons to select it. A button that is selected is shown with a filled circle; buttons that are not selected are shown with open circles.
Reference sequence	A sequence used as a prototype or benchmark sequences. If a sequence has been designated as Reference within Sequencher it has special properties.
Refrigerator	A special folder in which to store subsets of sequences which need to remain in your project.
SCF	Standard Chromatogram Format. A format used to store analyzed data or fragment data for a single sample.
Secondary peak	A chromatogram peak whose height may be less than that of the primary peak and whose presence may indicate that a heterozygote exists at that position.
Sequence	A small piece of DNA, decoded into its component bases, in order, represented as text. Sometimes in reverse order, depending upon the vagaries of the sequencing procedure.
Sequence Editor	The window used for editing DNA sequences. This editor looks much like a word processor and supports the standard text editing commands: Cut, Copy, Paste and Undo.
Slider	An on screen control that simulates the action of a slide control that might normally be used as a volume or dimmer control on a household appliance.
Title bar	The region at the top of a window on the computer.
VecBase	A cloning vector sequence database prepared by Friedhelm Pfeiffer, in collaboration with William Gilbert. Also a file with this name, in VecBase (CODATA) format, Part of the Sequencher distribution.